



Published in final edited form as:

*J Am Acad Audiol.* 2014 October ; 25(9): 869–892. doi:10.3766/jaaa.25.9.9.

## Non-native Listeners' Recognition of High-Variability Speech Using PRESTO

Terrin N. Tamati<sup>\*</sup> and David B. Pisoni<sup>\*,†,‡</sup>

<sup>\*</sup>Department of Linguistics, Indiana University, Bloomington, IN

<sup>†</sup>Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN

<sup>‡</sup>Department of Otolaryngology – Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN

### Abstract

**Background**—Natural variability in speech is a significant challenge to robust successful spoken word recognition. In everyday listening environments, listeners must quickly adapt and adjust to multiple sources of variability in both the signal and listening environments. High-variability speech may be particularly difficult to understand for non-native listeners, who have less experience with the second language (L2) phonological system and less detailed knowledge of sociolinguistic variation of the L2.

**Purpose**—The purpose of this study was to investigate the effects of high-variability sentences on non-native speech recognition and to explore the underlying sources of individual differences in speech recognition abilities of non-native listeners.

**Research Design**—Participants completed two sentence recognition tasks involving high-variability and low-variability sentences. They also completed a battery of behavioral tasks and self-report questionnaires designed to assess their indexical processing skills, vocabulary knowledge, and several core neurocognitive abilities.

**Study Sample**—Native speakers of Mandarin ( $n = 25$ ) living in the United States recruited from the Indiana University community participated in the current study. A native comparison group consisted of scores obtained from native speakers of English ( $n = 21$ ) in the Indiana University community taken from an earlier study.

**Data Collection and Analysis**—Speech recognition in high-variability listening conditions was assessed with a sentence recognition task using sentences from PRESTO (Perceptually Robust English Sentence Test Open-Set) mixed in 6-talker multitalker babble. Speech recognition in low-variability listening conditions was assessed using sentences from HINT (Hearing In Noise Test) mixed in 6-talker multitalker babble. Indexical processing skills were measured using a talker discrimination task, a gender discrimination task, and a forced-choice regional dialect categorization task. Vocabulary knowledge was assessed with the WordFam word familiarity test,

---

Terrin N. Tamati, Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405; ttamati@indiana.edu.

Portions of the research were presented at the Conference on Sources of Individual Linguistic Differences in Ottawa, Canada, March 2–4, 2012; and at the 41st Annual Conference of New Ways of Analyzing Variation in Bloomington, IN, October 25–28, 2012.

and executive functioning was assessed with the BRIEF-A (Behavioral Rating Inventory of Executive Function – Adult Version) self-report questionnaire. Scores from the non-native listeners on behavioral tasks and self-report questionnaires were compared with scores obtained from native listeners tested in a previous study and were examined for individual differences.

**Results**—Non-native keyword recognition scores were significantly lower on PRESTO sentences than on HINT sentences. Non-native listeners' keyword recognition scores were also lower than native listeners' scores on both sentence recognition tasks. Differences in performance on the sentence recognition tasks between non-native and native listeners were larger on PRESTO than on HINT, although group differences varied by signal-to-noise ratio. The non-native and native groups also differed in the ability to categorize talkers by region of origin and in vocabulary knowledge. Individual non-native word recognition accuracy on PRESTO sentences in multitalker babble at more favorable signal-to-noise ratios was found to be related to several BRIEF-A subscales and composite scores. However, non-native performance on PRESTO was not related to regional dialect categorization, talker and gender discrimination, or vocabulary knowledge.

**Conclusions**—High-variability sentences in multitalker babble were particularly challenging for non-native listeners. Difficulty under high-variability testing conditions was related to lack of experience with the L2, especially L2 sociolinguistic information, compared with native listeners. Individual differences among the non-native listeners were related to weaknesses in core neurocognitive abilities affecting behavioral control in everyday life.

## Keywords

Non-native; speech recognition; indexical variability; individual differences

## INTRODUCTION

In everyday situations, listeners will often encounter difficult listening conditions that impede the successful understanding of the talker's intended message. Although adverse listening conditions can involve background noise or competition from other talkers, natural variability in the speech signal also contributes to speech recognition difficulties (e.g., Pisoni, 1997; Mattys et al, 2012). Listeners make use of previous knowledge and experience to rapidly adapt to talker- and group-related variability encoded in the speech signal. However, even native listeners differ substantially in their ability to recognize speech in such conditions (Gilbert et al, 2013). In addition, these conditions may be particularly challenging to non-native listeners who have imperfect knowledge of their second language (L2) and less experience with sociolinguistic variability in the L2. Although a great deal of research on non-native speech recognition in adverse listening conditions has focused on recognizing speech in different types of background noise and/or competition (see García Lecumberri et al, 2010), little is still known about the effects of indexical variability in speech on non-native spoken word recognition, as well as the great diversity in basic speech recognition abilities in an L2.

Many factors affect a listener's ability to reliably recognize and understand speech. Among the potential factors, speech perception performance on any given task depends on the *background* (e.g., type of noise and/or background competition, reverberation, signal-to-

noise ratio [SNR]), *target signal* (e.g., talker(s), linguistic material, content, *individual perceiver* (e.g., native language, linguistic background, cognitive resources), *task goals* (e.g., keyword recognition, isolated word recognition, true/false judgments), and the interactions among these factors (see Gilbert et al, 2013). Furthermore, each of these factors can potentially contribute to the difficulty of a particular listening condition, making it less than ideal. Thus, to better understand non-native speech recognition in adverse listening conditions, it is important to consider the type of noise and/or competition, the characteristics of the target materials, and the individuals carrying out the task (Jenkins, 1979). The current study was designed to investigate non-native spoken word recognition in adverse listening conditions, particularly in high-variability listening conditions where large individual differences would be expected to emerge.

## Background Competition

In the real world outside of the clinic or research laboratory, listeners communicate in a variety of environments involving different types and levels of background noise and/or competition from other talkers. These everyday conditions differ from highly controlled laboratory settings, which routinely test the listener in ideal, quiet listening conditions, and pose significant challenges to the listener. Indeed, real-world degraded listening conditions have been found to be more difficult for normal-hearing native listeners (e.g., Gilbert et al, 2013), and are especially difficult for children (e.g., Elliott, 1979; Neuman and Hochberg, 1983; Fallon et al, 2000), hearing-impaired children and adults (e.g., Finitzo-Hieber and Tillman, 1978; Plomp, 1994), older adults (e.g., Humes and Roberts, 1990; Gordon-Salant and Fitzgibbons, 1995a; Gordon-Salant and Fitzgibbons, 1995b), and non-native listeners (e.g., Nábelek and Donahue, 1984; Takata and Nábelek, 1990). To a large degree, successful robust spoken word recognition depends on the type and level of the speech signals in the listening environment. Performance on speech recognition tasks is affected by spectral overlap from energetic masking as well as perceptual interference due to informational masking (e.g., Pollack, 1975; Brungart, 2001; Brungart et al, 2001; Calandruccio et al, 2010).

Competing speech signals are particularly challenging for the listener (e.g., Carhart et al, 1969; Carhart et al, 1975) because they introduce both energetic and informational masking into the listening environment (Van Engen and Bradlow, 2007; Helfer and Freyman, 2009). Background noise and/or competition can also interact with the target signal to influence the relative ease of speech recognition in a particular environment. Speech recognition is much poorer when target talkers and competing talkers share similar voice characteristics (Brungart, 2001; Brungart et al, 2001). In addition, regional dialect has a greater effect on speech recognition in more challenging SNRs (Clopper and Bradlow, 2008). Thus, considered either separately or together, background competition and target variability each play a role in determining a listener's ability to reliably recognize speech in a given environment.

Many studies of non-native speech perception have examined the effects of noise on speech recognition performance using a range of tasks and methodologies. Previous research has shown that background noise has a greater effect on speech recognition for non-native

listeners than native listeners when substantial linguistic and context information are available (for a recent review of studies on effects of noise on non-native speech perception, see García Lecumberri et al, 2010). Non-native listeners perform worse at poor SNRs than do native listeners (e.g., Black and Hast, 1962; Gat and Keith, 1978; Florentine et al, 1984; Florentine, 1985; Buus et al, 1986; Mayo et al, 1997; Meador et al, 2000; van Wijngaarden et al, 2002; von Hapsburg et al, 2004; García Lecumberri and Cooke, 2006). However, the differences in performance depend on the type of information-processing task and specific test materials. When the use of linguistic and/or context information in the target speech is limited or blocked, the effect of noise on speech perception is roughly the same for native and non-native listeners (e.g., Flege and Liu, 2001; Cutler et al, 2004; Rogers et al, 2006). Taken together, previous studies have suggested that background noise disproportionately affects speech perception performance by non-native listeners compared with native listeners, particularly for tasks that require non-native listeners to use contextual information and/or L2 phonological and lexical knowledge to understand meaningful speech.

### The Target Signal

In the real world, listeners must communicate with talkers of different developmental, linguistic, and social backgrounds. Along with the symbolic linguistic content of the utterance, highly detailed contextual information about specific talkers and social groups is also encoded in the speech signal (Abercrombie, 1967). Indexical variability related to talker-specific and group patterns conveys valuable information about the talker, such as his or her gender, age, region of origin, native language, and physical and mental states. The challenge for the listener is to simultaneously process both the linguistic and indexical information to recover the talker's intended message, and make use of the indexical information to draw inferences about the talker's vocal source characteristics (e.g., Nygaard, 2008). Indeed, previous studies have demonstrated that indexical variability encoded in the speech signal plays an important role in speech perception and spoken word recognition processes (e.g., Johnson and Mullennix, 1997; Pisoni, 1997; Cleary et al, 2005), because information about the vocal source is implicitly and automatically encoded into memory (e.g., Palmeri et al, 1993; Nygaard et al, 1995) and later used to make inferences about the talker in a range of communicative environments (e.g., Ptacek and Sander, 1966; Lass et al, 1976; Van Lancker et al, 1985a; Van Lancker et al, 1985b; Van Bezooijen and Gooskens, 1999; Clopper and Pisoni, 2004b; Kreiman and Van Lancker Sidtis, 2011). Listeners can also learn to use talker-specific voice patterns in the signal to facilitate speech recognition in degraded listening conditions (e.g., Nygaard et al, 1994; Nygaard and Pisoni, 1998). In addition, a recent study investigating the neural mechanisms involved in recognizing speech from different talkers has provided further evidence that indexical and linguistic information are processed automatically and simultaneously in speech recognition (von Kriegstein et al, 2010).

While listeners are often able to adapt to and use this “extralinguistic” information in the speech signal, indexical variability can also impede successful speech recognition. Some individual talkers are inherently more intelligible than others (Bradlow et al, 1996). Native listeners also have more difficulty understanding foreign-accented speech (e.g., Lane, 1963), as well as speech from talkers who come from unfamiliar or marked dialect regions (e.g.,

Mason, 1946; Labov and Ash, 1997; Clopper and Bradlow, 2008). The number and types of different talkers within a task also influence speech recognition. Understanding speech from multiple talkers in a speech recognition test is more difficult than understanding speech produced by only a single talker (Peters, 1955; Creelman, 1957; Mullenix et al, 1989). Use of multiple talkers introduces variability in indexical information and listeners must rapidly adjust and adapt to these changing conditions, which often prevent or reduce the learning of talker-contingent details (Nygaard et al, 1994; Nygaard and Pisoni, 1998).

A number of studies have reported that non-native listeners are highly sensitive to variability in the speech signal and are capable of accommodating to and learning new patterns of indexical variation in an L2 (e.g., Clopper and Bradlow, 2009; Ikeno and Hansen, 2007). Several studies have also investigated the ability of non-native listeners to recognize speech produced by multiple talkers from different regions of origin in their L2. Cooke et al (2008) examined native Spanish speakers' recognition of keywords in English sentences produced by multiple talkers, with a variety of accents, in different noise and competing speech conditions. In their study, although non-native listeners made more errors than native listeners overall, both groups performed similarly by gender; female talkers were more intelligible for all listeners. However, the intelligibility of individual talkers varied for the native and non-native listeners in quiet and more favorable noise conditions. Cooke et al (2008) suggested that some of these difficulties may have been the result of dialectal variation, because non-native listeners had particular difficulty with the nonstandard Scottish variety with which they lacked exposure and familiarity.

Fox and McGory (2007) also found that native Japanese learners of English (as well as native English speakers), regardless of location of residence (Ohio or Alabama), were more accurate in identifying vowels produced by a General American English talker than a marked Southern American English talker. However, unlike native listeners from Alabama, who were also quite accurate for the same set of Southern vowels, Japanese listeners living in Alabama did not exhibit a processing benefit for the Southern variety. Similarly, Eisenstein and Verdi (1985) found that English learners in New York City had much more difficulty in understanding African American English than either New York English or General American English. The findings from these studies suggest that although non-native listeners can use and benefit from low-level acoustic-phonetic talker differences related to gender, lack of exposure to different accents in the L2 may affect a listener's ability to understand speech produced by talkers from different dialect regions. While non-native listeners show benefits for a standard variety of the L2, they have more difficulty understanding nonstandard or unfamiliar varieties, even after substantial contact and exposure to these speaking styles.

### **The Individual Perceiver**

Groups and individuals also vary substantially in their ability to understand speech. A listener's unique developmental linguistic history influences his or her ability to recognize speech. The language backgrounds of non-native listeners have consistently been shown to influence L2 speech perception abilities. These factors include first language (e.g., MacKay et al, 2001a; MacKay et al, 2001b; García Lecumberri and Cooke, 2006), age and length of

exposure to the L2 (e.g., Oyama, 1976; MacKay et al, 2001a; MacKay et al, 2001b; Ingvalson et al, 2011), and frequency and amount of L2 use (e.g., Oyama, 1976; Flege et al, 1997; Flege and MacKay, 2004; Ingvalson et al, 2011). These group-related factors have also been found to affect a non-native listener's ability to understand test materials including target variability, such as foreign-accented speech (e.g., Bent and Bradlow, 2003; Imai et al, 2005). Even within a native listener population, the linguistic and developmental history of an individual listener has been found to influence his or her speech recognition performance. The effect of early linguistic experience interacts with the indexical characteristics of the target materials produced by different talkers.

Previous research has found that more exposure and experience facilitate speech recognition. For example, the speech of deaf children is more intelligible to listeners who are more familiar with deaf talkers than naïve inexperienced listeners (McGarr, 1983). Similarly, listeners are better able to understand talkers from familiar local or supraregional dialects than unfamiliar or novel dialect regions (e.g., Mason, 1946; Labov and Ash, 1997; Clopper and Bradlow, 2008). Taken together, these studies consistently demonstrate that prior linguistic experience can influence individual performance on a wide range of speech recognition tasks.

Beyond group factors, several studies have also demonstrated large individual differences in speech recognition abilities of young monolingual listeners. Recently, Gilbert et al (2013) examined speech recognition with high-variability sentences that included diverse indexical and linguistics characteristics. They found that young college-aged, normal-hearing native listeners varied substantially in their ability to recognize high-variability sentences in 6-talker multitalker babble across four SNRs. The sentence materials used in the Gilbert et al (2013) study came from PRESTO (Perceptually Robust English Sentence Test Open-set), a new high-variability sentence recognition test that contains variability in talker, gender, and regional dialect. Participants' (N = 121) keyword accuracy scores ranged from approximately 40–76% across four SNRs, suggesting that although all listeners had hearing thresholds within normal limits, some listeners were better than others at recognizing highly variable speech in multitalker babble. Other studies have also reported substantial individual variability in speech recognition abilities in a variety of listening environments and conditions (e.g., Neff and Dethlefs, 1995; Richards and Zeng, 2001; Wightman et al, 2010).

At the present time, however, little is known about the underlying sensory and neurocognitive factors that influence speech recognition in noise. Several studies have suggested that individual differences in a small number of core neurocognitive abilities affect speech recognition performance. Research on pediatric cochlear implant users has shown that individual speech recognition abilities and language outcomes measured later in life are related to several core neurocognitive information-processing abilities, such as short-term and working memory capacity, verbal rehearsal speed, and scanning/retrieval processes (Pisoni and Geers, 2000; Cleary et al, 2000; Pisoni and Cleary, 2003; Pisoni and Cleary, 2003; Pisoni et al, 2011). Similarly, Beer et al (2011) found that parent reports of working memory strengths and weaknesses on the BRIEF behavioral rating questionnaire correlated significantly with speech perception in noise in children with cochlear implants. In adults, attentional skills (Jesse and Janse, 2012) and implicit learning abilities (Conway et al, 2010)



have also been found to be related to performance on speech perception tasks. Neurocognitive abilities of older adults have also been found to play an important role in speech recognition, especially when auditory factors are accounted for (e.g., van Rooij and Plomp, 1990; Gordon-Salant and Fitzgibbons, 1997; Humes, 2007; Akeroyd, 2008).

In a recent study, Tamati et al (2013) examined differences in the indexical processing skills and neurocognitive abilities between good and poor listeners on PRESTO sentences used in an earlier study (Gilbert et al, 2013). We found that good listeners on PRESTO were better at discriminating talkers by gender and categorizing talkers by regional of origin (Tamati et al, 2013). In addition, we found that good listeners had greater short-term and working memory capacity and more knowledge of English vocabulary, as well as better executive functioning skills in domains related to cognitive load. These findings suggest that individual differences in speech recognition on high-variability sentences may be closely related to the ability to perceive and encode detailed episodic information in the speech signal. Listeners benefit from more robust, highly detailed lexical representations and memory codes in long-term memory for both signal and context information in high-variability sentence recognition tasks.

Several studies have also explored individual differences in L2 abilities. Even in groups of non-native listeners who have similar native language backgrounds and experiences, individual learners tend to vary greatly in their speech perception abilities in the L2. Neurocognitive abilities of these listeners may also underlie individual differences in L2 phonological development. Recently, Darcy et al (2012) assessed individual differences in phonological processing skills in an L2 and the contribution of neurocognitive abilities to performance on several conventional L2 phonological processing tasks. Although the relationship between neurocognitive abilities and speech perception performances varied from task to task, Darcy et al (2012) found evidence that better working memory capacity, processing speed, lexical retrieval, and executive function were related to better speech perception performance. Other studies have also provided converging evidence that individual differences in core neurocognitive abilities may influence L2 acquisition and speech perception (e.g., Salthouse, 1996; Segalowitz, 1997; Miyake and Friedman, 1998). Taken together, studies on both native and non-native speech perception have suggested close links between individual differences in speech recognition abilities and several underlying core neurocognitive processes related to the processing operations used to encode, store, and retrieve phonological and lexical representations of speech.

## THE CURRENT STUDY

The current study investigated non-native listeners' speech recognition ability on high-variability English sentences. Non-native speakers of American English residing in the United States (U.S.) completed a battery of speech perception and indexical processing tasks, along with several self-report questionnaires. To reduce the number of potential variables influencing speech recognition abilities in the L2, all non-native listeners shared the same first language (Mandarin); the amount and types of exposure to English in the L2 setting were also controlled.

To examine the contributions of different types of variability in the speech signal to speech recognition in challenging listening conditions, listeners completed both a high-variability and a low-variability sentence recognition task in multitalker babble. Listeners also completed several indexical processing tasks to assess their ability to perceive indexical/extralinguistic attributes of speech in American English. In addition to the indexical processing tasks, self-report questionnaires of language experience, neurocognitive abilities, and vocabulary knowledge were used to uncover additional factors that might influence a non-native listener's ability to successfully recognize high-variability speech produced by multiple talkers. Thus, the goal of the present study was to examine the effects of variability in the speech signal on speech recognition in an L2, the processing of indexical information in an L2, and factors contributing to individual differences in speech recognition in highly variable listening conditions.

### Speech Recognition Abilities

Speech recognition abilities were assessed with two-sentence recognition tasks in multitalker babble that differed in the amount of variability in the stimulus materials. The high-variability task was constructed using sentences from PRESTO (Perceptually Robust English Sentence Test Open-Set; see Gilbert et al, 2013). The PRESTO test contains both linguistic variability in words and sentences and indexical variability in talker, gender, and regional dialect. Thus, these high-variability test materials allowed us to examine how non-native listeners rapidly adjust to multiple sources of variability in the speech signal to successfully recognize keywords in meaningful sentences. Although not truly representative of everyday listening situations, the high-variability PRESTO sentences in multitalker babble should provide new insights into how L2 listeners deal with the challenges of real-word listening conditions where they have to understand multiple talkers from diverse backgrounds against competing background talkers.

The low-variability task was created with sentences selected from the original version of the Hearing In Noise Test (HINT) that includes lists of 10 low-variability sentences (Nilsson et al, 1994). Unlike the high-variability PRESTO sentences, the HINT sentences include very little linguistic variability and include no variability in talker, gender, and regional dialect because all sentences are produced by a single male talker. Diverging from the original HINT and conventional protocol, the HINT sentences in the current study were presented in multitalker babble and responses were scored by keywords correct. Using the same type and levels of competition and the same methods of scoring allowed for a direct comparison in keyword recognition between low-variability HINT listening conditions and the high-variability PRESTO listening conditions.

Given the previous studies that have shown that non-native listeners have much more difficulty recognizing speech in a variety of listening conditions and environments than native listeners when higher-level linguistic and context information is available to the listener, we predicted that non-native listeners would be less accurate at recognizing keywords than native listeners on both types of test sentences (PRESTO and HINT). Keyword accuracy on HINT sentences was also expected to be better than PRESTO because these materials contain little linguistic and indexical variability. Additionally, given our



listeners' lack of experience with American English, and specifically, exposure to regional dialect variability, we predicted that the PRESTO sentences would be more difficult than HINT sentences for the non-native listeners compared with a group of native listeners.

### Processing of Indexical Information in Speech

Three perceptual tasks were used to assess the processing of indexical information by non-native listeners. Listeners completed a gender discrimination task to assess their ability to rapidly discriminate between male and female talkers and a talker discrimination task to assess their ability to rapidly discriminate within-gender talker differences. Additionally, all listeners completed a forced-choice regional dialect categorization task to assess the listeners' ability to perceive and classify dialect-specific differences and use knowledge of these differences to identify an unfamiliar talker's region of origin. In addition to encoding important sources of indexical information in everyday speech communication, acoustic cues specifying gender, talker, and regional dialect are all key attributes that vary in the PRESTO sentences materials. The non-native listeners were not expected to have any difficulty in either the talker or gender discrimination tasks because these tasks both involve discrimination of large acoustic-phonetic differences in the speech signal that encode talker and gender information, which are highly discriminable to native speakers and do not rely on linguistic knowledge or categorization skills.

With respect to non-native listeners' ability to categorize talkers by region of origin, previous studies have found that non-native listeners are sensitive to dialect-specific details in speech, but they are not as accurate as native listeners in categorization tasks and likely rely on different attributes of the signals in making their dialect classification decisions. For example, Clopper and Bradlow (2009) found that although non-native listeners were fairly accurate in grouping talkers by regional dialect using an auditory-free classification task in their L2 (English), native speakers were more accurate overall. Examining the two groups' performance, Clopper and Bradlow (2009) also found that non-native listeners relied more heavily on subtle acoustic properties of the test signals that the non-native listeners were unable to use, such as knowledge of how different features group together to index dialects, especially when there was much within-dialect variability. The forced-choice regional dialect categorization task used in the present study should be even more challenging for the non-native listeners because this task requires the use of previous knowledge of, and experience with, regional dialect variation in American English to learn the dialect-specific acoustic-phonetic cues relevant for explicit dialect categorization of American English. Thus, non-native listeners were expected to perform more poorly on the forced-choice regional dialect categorization task than native listeners.

On the basis of previous research reported by Tamati et al (2013), we expected that individual differences in the ability of non-native listeners to encode highly detailed episodic information in speech would be related to their speech recognition abilities on PRESTO. Thus, we predicted that non-native listeners who are better able to encode detailed indexical information in speech would be better at discriminating talker and gender information and in categorizing talkers by region of origin, and would therefore be better at recognizing PRESTO sentences in multitalker babble.

## Other Factors Contributing to Performance

The contribution of vocabulary knowledge to group and individual differences in the ability to recognize keywords in highly variable sentences was also investigated. Lexicon size and lexical connectivity have consistently been found to play important foundational roles in speech perception and spoken word recognition (e.g., Ganong, 1980; Pisoni et al, 1985; Samuel, 1986; Altieri et al, 2010). Listeners are able to use lexical knowledge to recognize words from partial and degraded information (e.g., Pisoni et al, 1985; Altieri et al, 2010). Non-native listeners who have less experience with English words have more difficulty using their downstream lexical knowledge to help resolve ambiguous or compromised acoustic-phonetic information to arrive at the correct lexical item (e.g., Bradlow and Pisoni, 1999; Ezzatian et al, 2010; Bundgaard-Nielsen et al, 2011).

Tamati et al (2013) also found that listeners who had more difficulty in recognizing PRESTO sentences in multitalker babble also knew fewer words than listeners who were better at recognizing words on PRESTO. Differences in vocabulary knowledge among the non-native listeners were also expected to contribute to performance variability in recognizing words on PRESTO. Non-native listeners with large L2 vocabulary sizes would be able to use their knowledge of English words and phonotactics to recognize more words in adverse listening conditions. Thus, although we expected that the non-native listeners would be less familiar with English words overall than native listeners, within the non-native listener group, we also predicted that listeners who know more English words would also perform better on PRESTO.

As reported in previous studies, differences in several core neurocognitive processes were expected to be related to speech recognition performance. In particular, earlier studies have found that measures of executive functioning and cognitive control were related to individual differences in speech recognition skills (Beer et al, 2011; Tamati et al, 2013). To examine the contribution of executive function and cognitive control processes, we assessed real-world everyday executive functioning skills among the non-native listeners in this study using the Behavioral Rating Inventory of Executive Function – Adult Version, a self-report behavior-rating scale that is used to assess executive functioning (BRIEF-A, Psychological Assessment Resources, Roth et al, 2005). Executive functioning is particularly relevant to speech recognition in adverse listening conditions because it reflects the operation of the cognitive control system that supervises and manages several core cognitive processes used in real-time spoken language comprehension. Executive functioning controls processes such as attention and inhibition, working memory, and behavior regulation, and as such, it is thought to play a foundational role in many everyday activities and behaviors, determining an individual's ability to initiate and terminate behaviors, appropriately adjust to new and changing situations, and control attentional and processing resources (Barkley, 1997, 2012). Non-native listeners who show better executive functioning skills on the BRIEF-A self-report questionnaire were expected to be better at recognizing keywords in PRESTO sentences, which include novel sentences that are produced by different talkers that change from trial to trial. To recognize spoken words robustly in these challenging conditions, the listener must rapidly adjust and adapt to new talkers and novel regional dialects on each trial of the task.

The present study reduced the potential contributions of several background factors to speech recognition abilities by selecting only non-native listeners with the same native language, similar educational history, and comparable lengths (and locations) of residency in the U.S. However, given the well-documented contributions of language background and experience to speech recognition performance, some factors, including length of residency, age of arrival in the U.S., and age of initial English instruction, were taken into account in assessing individual L2 listeners' speech recognition abilities. Given that the exposure of the participants to English was limited by design, we expected that there would not be enough variability in their language backgrounds to significantly contribute to performance. Although a lack of relationship would not mean that these factors do not play any role in speech recognition in high-variability listening conditions, limiting the variability in language background provides a way to constrain other contributing factors in order to examine the effects of indexical processing skills and neurocognitive abilities independently from demographic factors.

To summarize, the present study had three primary goals: (1) to assess the effect of high-variability stimulus materials on sentence recognition abilities of non-native listeners; (2) to investigate non-native listeners' perception of indexical information in the speech signal; and (3) to examine individual differences among non-native listeners in the processing of indexical information in speech, executive functioning, and vocabulary knowledge to determine if these factors are related to the ability to understand highly variable speech. Non-native listeners were expected to perform worse than native listeners on sentence recognition using both the high-variability PRESTO sentences and the low-variability HINT sentences, but we predicted that they would experience more difficulty on the PRESTO sentences than on the HINT sentences compared with the native listeners. Within the non-native group, good PRESTO non-native listeners were expected to show more accurate and/or faster processing of indexical information in American English, better executive functioning, and a greater lexical knowledge of American English. Thus, individual differences in basic sensory, perceptual, and neurocognitive abilities were predicted to underlie speech recognition abilities in highly variable, adverse listening conditions in an L2.

## METHODS

### Participants

A total of 28 native speakers of Mandarin Chinese participated in the current study. All of the participants were students at Indiana University in Bloomington, IN. Before analysis, three of the participants were excluded because they were unable to complete all of the tasks in the study session. The remaining 25 participants (22 females and 3 males) were all native speakers of Mandarin from main land China, ranging in age from 19–34 yr. Length of residency in the U.S. ranged from roughly 1–27 mo. Of the 28 participants, 23 had only lived in Bloomington, IN, while in the U.S. The other two participants lived in Bloomington, IN, and only one other location where a General American dialect variety (the variety spoken in Bloomington, IN) was spoken for less than 1 yr. Non-native participants were highly educated undergraduate and graduate students, having met the criteria for admittance

to Indiana University. None of the participants reported a significant history of hearing loss or speech disorders at the time of testing. They received \$20 for 120 min of participation as compensation.

For the purpose of comparison, data from the non-native listener group were compared with scores obtained from a group of native listeners who had participated in a previous study in which they completed the same tasks used in the current study, along with other perceptual and neurocognitive tasks (see Gilbert et al, 2013; Tamati et al, 2013). The native comparison group consisted of native speakers of American English who fell in the lower quartile of recognition accuracy obtained from a distribution of 121 native listeners on PRESTO in multitalker babble (LoPRESTO group; Gilbert et al, 2013). As such, this group provided a conservative estimate of average native ability to understand highly variable speech on PRESTO. Given that patterns of performance on PRESTO and HINT in Gilbert et al (2013) were similar for all groups tested, we believed that although these listeners obtained lower-than-average scores on PRESTO (compared with other college-aged, normal-hearing young adults), they were representative of native recognition of highly variable speech by normal-hearing, typically developing, and highly educated native speakers of American English. The native group included 21 young adults (14 females and 7 males), ranging in age from 18–24 yr. All participants were normal-hearing native speakers of American English who had reported no significant history of hearing loss or speech disorders at the time of testing. (For a more detailed description of the native/LoPRESTO listener group, see Gilbert et al, 2013; Tamati et al, 2013.)

## Materials and Procedures

Participants in the present study were tested in groups of four or fewer in a quiet room, where they completed a series of computer-based perceptual tasks and self-report questionnaires. At the end of testing, participants also completed a short questionnaire on their language background and residential history. For the computer-based tasks, each participant was seated in an enclosed testing carrel in front of a Power Mac G4 computer running Mac OS 9.2 with diotic output to Beyer Dynamic DT-100 circumaural headphones. Computer-based experimental tasks were controlled by custom PsyScript 5.1d3 scripts. Each test sentence .wav file was equated to the same root-mean-square amplitude level. Output levels of the test sentences for all computer based perception tasks were calibrated to be approximately 64 dB SPL. All stimulus items in the perception tasks were presented diotically to the participants. Procedures for individual tasks are described below.

### Speech Recognition Tasks

**High-Variability Listening Conditions, PRESTO:** Speech recognition in high-variability listening conditions was assessed using 10 PRESTO sentence lists mixed in multitalker babble. The PRESTO test consists of meaningful English sentences obtained from the TIMIT speech corpus (Garofolo et al, 1993), which includes sentences produced by multiple male and female talkers from eight different dialect regions of the U.S. Felty (2008) originally selected a subset of the TIMIT sentences to create sentences lists (18 sentences per list) that included nine different male and nine different female talkers from at least five different U.S. dialect regions. Each list contained 76 keywords and was balanced for average

word familiarity ( $M = 6.9$ ) and average log keyword frequency ( $M = 2.5$ ) across sentence lists (Nusbaum et al, 1984). Within each test list, no sentence was repeated. PRESTO sentences also differed in the number of keywords and overall length, as well as syntactic structure (see Gilbert et al, 2013). The PRESTO lists used in the current study included 180 utterances produced by 169 talkers (158 produced one sentence and 11 talkers produced two sentences). Talkers came from one of the following eight different U.S. dialect regions, as identified in the TIMIT database based on the geographic region(s) where they had spent their childhood: New England, Northern, North Midland, South Midland, Southern, New York City, Western, and Army Brat (moved around during childhood).

For each test sentence, digital audio files (equated in signal level) were mixed with random samples from a 5 min stream of 6-talker babble composed of three male and three female talkers. All talkers in the multitalker babble were from a region of the U.S. where General American is spoken (with General American operationally defined as a non marked, or suprasegmental, dialect, here identified as speakers from New England, North Midland, South Midland, or West). Four different SNRs were presented. Two PRESTO lists were presented at +3 dB SNR, three lists were presented at 0 dB SNR, three lists were presented at -3 dB SNR, and two were presented at -5 dB SNR. For presentation, the level of the target sentences was held constant and the level of the multitalker babble was varied. More lists were presented at 0 and -3 dB SNR because we predicted that ceiling or floor effects would be less likely to occur at these middle SNR levels.

**Low-Variability Listening Conditions, HINT:** Speech recognition in low-variability listening conditions was assessed using sentences from the HINT (Nilsson et al, 1994) in multitalker babble. The HINT test is a commonly used, single-talker sentence recognition test for American English based on British English Bamford-Kowal-Bench materials originally created by Bench et al (1979). Four HINT lists, lists 1–4, were used for a total of 40 sentences. All sentences were produced by a single male talker with speech that is characteristic of General American English. HINT sentences were 4–7 words in length. Sentences were originally selected to be equally intelligible at a fixed-noise level as scored by sentence. However, the presentation of these sentences, when equated for root-mean-square amplitude and the non conventional keyword scoring, may reduce the equality of the intelligibility of the sentences and alter list equivalency. Distribution of phonemes was also balanced across lists of 10 sentences. The sentences were mixed with multitalker babble using the same methods as the PRESTO sentences, although this is not the conventional method of administering or scoring the HINT test (see Nilsson et al, 1994). One HINT list was presented at each of the four SNRs (+3, 0, -3, and -5 dB SNR).

For both tests, sentences were randomly presented binaurally to the listeners over headphones. Randomized presentation of target sentences and SNRs was used to maintain attention and effort and reduce fatigue. Thus, variability in the test sentence and background competition level varied from trial to trial, and each participant received a different random order of presentation. Listeners were asked to carefully listen to each sentence and type in all of the words they had heard into a dialog box displayed on a computer monitor. They were encouraged to give partial answers or guesses if they were unsure of their responses. The listeners heard each sentence only once. The experiment was self-paced. Participants

could take as much time as they liked to respond before they moved on to the next trial when they were ready. Responses for both PRESTO and HINT sentences were scored off-line by hand for keywords correct. Given that the PRESTO test was developed to be scored by keywords (Felty, 2008), for the purpose of the current study, only the words designated as keywords were scored. To attempt to obtain similar scoring procedures for the HINT sentences, which were not originally designed to be scored by keyword, keywords were selected from the set of HINT sentences before testing. To obtain a set of HINT keywords similar to the PRESTO keywords, all content words in the sentence—including nouns, verbs, adjectives, adverbs, and pronouns—were selected. Other function words, including articles, prepositions, and auxiliary verbs, were excluded. Correct morphological endings were required. Homophones and responses with minor spelling errors were counted as correct responses.

Before statistical analysis, in order to account for the possible impact of ceiling or floor effects, we transformed proportion-correct accuracy scores to rationalized arcsine units (Studebaker, 1985). For ease of interpretation and clarity, all reported values are percent-correct keyword accuracy scores.

### Indexical Processing Tasks

**Gender Discrimination:** Participants completed a gender discrimination task. This task assessed the participants' ability to perceive gender-specific information in isolated sentences. Four talkers from the Indiana Multitalker Sentence Database (Bradlow et al, 1996; Karl and Pisoni, 1994) were selected for this task. Half of the talkers were female and half were male. Materials consisted of 32 unique utterances, 8 for each of four talkers (two female and two male). On each trial, listeners were presented with a pair of sentences, separated by 1000 msec of silence. Participants were asked to decide if the talkers in each pair of sentences were the same or different genders. Overall, all talkers were paired with themselves once and all other talkers six times, for eight presentations per talker, with no utterance repeated. Thus, the entire task included a total of 16 trials, of which 8 pairs were "Same-Gender" talkers and 8 pairs were "Different-Gender" talkers. Participants were instructed to respond as quickly as possible without compromising accuracy. They responded by pressing one of two buttons on a response box. For all participants, "Same-Gender" responses were always represented by the button farthest to the right, and "Different-Gender" responses always the farthest button to the left. Response accuracy (correctly responding "Same" to a "Same-Gender" trial or "Different" to a "Different-Gender" trial) and response times (RTs) were collected and analyzed separately.

**Talker Discrimination:** A talker discrimination task was used to assess the participants' ability to perceive within-gender talker-specific differences. The talker discrimination task was based on the methodology originally developed by Cleary and Pisoni (2002). Six different talkers from the Indiana Multitalker Sentence Database (Bradlow et al, 1996; Karl and Pisoni, 1994) were selected for this task. Half of the talkers were female and half were male. Materials consisted of 48 unique utterances, 8 for each of six talkers (three female and three male). On each trial, listeners were presented with a pair of sentences, separated by 1000 msec of silence, that were produced by a single talker or by two different talkers.



Listeners were asked to decide if the two sentences were produced by the same talker or by two different talkers. Sentences produced by male talkers were always paired with another sentence produced by the same or a different male talker; similarly, female talkers were always paired together. Overall, each talker was paired twice with a different (but same gender) talker, and four times with themselves, for a total of six presentations per talker. No utterance was repeated. The talker discrimination task consisted of a total of 24 trials, of which 12 were “Same-Talker” trials and 12 were “Different-Talker” trials.

Participants were instructed to respond as quickly as possible without compromising accuracy. They responded by pressing one of two buttons on a button box. For all participants, “Same-Talker” responses were always represented by the button farthest to the right, and “Different-Talker” responses always the farthest button to the left. Accuracy (correctly responding “Same” to a “Same-Talker” trial or responding “Different” to a “Different-Talker” trial) and RTs were collected and analyzed separately.

**Regional Dialect Categorization:** Participants completed a six-alternative forced-choice regional dialect categorization task. This task was used to assess a listener’s ability to perceive dialect-specific information in the speech signal and use stored knowledge of regional dialects to identify the region of origin of unfamiliar talkers. The forced-choice dialect categorization task used in this study was based on the methodology originally developed by Clopper and Pisoni (2004b). The talkers and test sentences for this task were selected from the TIMIT acoustic-phonetic speech corpus (Garofolo et al, 1993). A total of 48 talkers, 24 female and 24 male, were used in this task. Eight talkers (four female and four male) were from each of the following six dialect regions: New England, North Midland, South Midland, North, South, and West. Two sentences were used for each talker for the task. One sentence, “She had your suit in greasy wash water all year,” was the same for all talkers. This sentence is one of the “baseline” calibration sentences collected from all talkers in the TIMIT database and was designed to obtain dialectal differences. The other sentence was unique to each talker and was selected to contain phonological/phonetic variation representative for each of the six dialect regions (Clopper and Pisoni, 2004b).

The dialect categorization task was divided into two blocks. In the first block, participants heard the talkers saying the same baseline sentence. In the second block, participants heard the talkers produce the unique sentences. Each talker was presented one time per block in a random order, for a total of 96 trials (48 trials during the first block and 48 trials during the second block). Participants were permitted to take a break after each block. On each trial, participants heard a single talker and were asked to select the region where they thought the talker was from using a closed set of six response alternatives. Participants were required to choose from the six dialect regions (New England, North Midland, South Midland, North, South, and West) represented on a colored graphical map of the U.S. displayed on a computer monitor. Participants entered their responses by clicking on a labeled box located within each dialect region; they could take as long as they wanted to respond on this task. Once they responded, the next trial began. Participants’ responses were collected and coded for the dialect region selected and overall accuracy. Both incorrect and correct responses were analyzed.

## Self-Report Questionnaire on Vocabulary Knowledge

**WordFam:** Participants completed the WordFam test (Pisoni, 2007), a self-report word familiarity rating questionnaire originally developed by Lewellen et al (1993) to study individual differences in lexical knowledge and organization. Responses on the WordFam test provide a measure of vocabulary knowledge. In this task, participants were instructed to rate how familiar they were with a set of English words using a 7-point scale, ranging from 1 (“You have never seen or heard the word before”) to 7 (“You recognize the word and are confident that you know the meaning of the word”). The WordFam questionnaire contained a total of 150 English words, including 50 low-familiarity, 50 medium-familiarity, and 50 high-familiarity items based on ratings obtained from Nusbaum et al (1984). Participants responded by marking the number corresponding to the familiarity rating for a given item. Responses for all 150 items were collected and averaged for each of the three lexical familiarity conditions.

## Self-Report Questionnaire on Executive Function

**BRIEF-A:** Participants also completed the BRIEF-A (Psychological Assessment Resources, Roth et al, 2005). The BRIEF-A is a self-report questionnaire designed to assess a participant’s rating of his or her own executive functions in everyday life. The BRIEF-A consists of 75 statements to which participants must respond if their behavior is a problem: *Never* (1 point), *Sometimes* (2 points), or *Often* (3 points). The BRIEF-A is used to evaluate nine clinical domains of executive function: Shift, Inhibit, Emotional Control, Self-Monitor, Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials. These nine domains are grouped into two aggregate indexes: a Behavioral Regulation Index (BRI) and a Metacognitive Index (MI), and an overall global General Executive Composite (GEC) score.

The BRI of the BRIEF-A consists of Shift, Inhibit, and Emotional Control scales. The Shift scale is related to the ability to move or change from one situation, topic, or task to another. The Inhibit scale is related to the ability to inhibit or stop oneself from acting on impulse in different situations. Emotional Control refers to one’s ability to control or modulate emotional behavior. Overall, the BRI is related to one’s ability to appropriately change or adapt emotional behavior demonstrating good inhibitory control.

The MI of the BRIEF-A consists of the Self-Monitoring, Initiate, Working Memory, Plan/Organize, Task Monitoring, and Organization of Materials scales. Self-Monitor refers to one’s ability to evaluate how his or her actions or behaviors affect others. Initiate is related to the ability to start a new task or generate new, independent thoughts or ideas. Working Memory scale assesses the participant’s ability to hold information in memory for completing, or taking the necessary steps to complete, a task or goal. Plan/Organize refers to the ability to manage information to complete current tasks or in anticipation of future tasks. Task Monitoring refers to one’s ability to evaluate his or her actions or behaviors during or after a task or activity. The Organization of Materials scale is related to the orderliness or organization of one’s belongings and space or actions. Overall, the MI is related to the ability to appropriately manage and complete different tasks and actions. The GEC of the

BRIEF-A includes all individual scales and provides an overall composite assessment of executive functioning in everyday life.

## RESULTS

### Group Differences: Non-native versus Native Comparison

#### Speech Recognition Tasks

**High-Variability Listening Conditions, PRESTO:** A mean keyword correct accuracy score on PRESTO was calculated separately for each SNR condition, and overall. Performance on PRESTO was very poor for the non-native listeners. Averaged across all SNRs, mean accuracy was 23.2% keywords correct ( $SD = 5.7\%$ ). However, performance of individual non-native subjects varied. Mean word recognition accuracy across all SNRs ranged from 12.0–35.3% within the non-native group. Figure 1 shows individual and group mean keyword accuracy scores on PRESTO for non-native listeners and the native listeners reproduced from Gilbert et al (2013). As expected, paired comparison  $t$ -tests revealed that accuracy was better at more favorable SNRs than less favorable SNRs (all pairwise comparisons were significant, all  $p < 0.001$ ). Strong positive correlations were also observed for all pairwise comparisons of the word recognition scores across all SNRs ( $r = 0.74$ – $0.87$ , all  $p < 0.001$ , two-tailed), suggesting that individual performance was highly consistent within individual subjects at different SNRs. Mean accuracy scores on PRESTO at each SNR and overall across all trials and SNRs are given in Table 1. Note that overall PRESTO and HINT scores should be compared with caution because they may be slightly affected by the distribution of trials by SNR, which differed for PRESTO and HINT (PRESTO, +3 dB SNR: 20%; 0 dB SNR: 30%; –3 dB SNR: 30%; –5 dB SNR: 20%; HINT, all SNRs: 25%). To aid in the comparison of overall performance on PRESTO and HINT, adjusted PRESTO (and HINT) scores were calculated from each listener's average of his or her four overall SNR scores, each of which then contributed to 25% of the adjusted score. Adjusted overall PRESTO scores were nearly identical to the overall scores across all trials for both non-native (mean adjusted overall PRESTO score = 23.3%) and native (mean adjusted overall PRESTO score = 55.6%) groups.

To compare non-native high-variability keyword recognition performance with sentence recognition scores of the native listeners, we carried out a repeated-measures ANOVA on the word recognition scores with listener group (non-native, native) as the between-subject factor and SNR as a within-subject factor. The ANOVA revealed significant main effects of SNR [ $F_{(3,132)} = 906.75$ ,  $p < 0.001$ ] and listener group [ $F_{(1,44)} = 463.58$ ,  $p < 0.001$ ], as well as a significant SNR  $\times$  listener group interaction [ $F_{(3,132)} = 73.21$ ,  $p < 0.001$ ]. The significant main effect of SNR confirms that all listeners performed better at more favorable SNRs, as reported above, and will not be further analyzed. To explore the main effect of listener group and the SNR  $\times$  listener group interaction, we carried out independent-samples  $t$ -tests between groups for each SNR. The non-native listeners performed more poorly than native listeners across all SNRs [ $t_{(44)} = 17.54$ – $20.08$ , all  $p < 0.001$ ]. Although the groups were significantly different in all SNR conditions, the difference between the groups varied by SNR, and was larger at the moderate SNRs. Figure 2 displays the average keyword accuracy on PRESTO by SNR for both the non-native and native listener groups.

**Low-Variability Listening Conditions, HINT:** Mean accuracy scores were also calculated for the HINT sentences. Averaged across all four SNRs, mean accuracy for the non-native listeners was 38.5% (SD = 7.4%). However, individual non-native listeners' performance also varied. Mean word recognition accuracy across all trials and all SNRs ranged from 27.6–50.0% within the non-native group. Paired-comparison *t*-tests revealed that performance was better at more favorable SNRs than less favorable SNRs (all pairwise comparisons were significant, all  $p < 0.001$ ). Strong positive correlations were also observed for some, but not all, pairwise comparisons of the word recognition scores across SNRs. Performance at +3 dB SNR correlated positively with performance at 0 dB SNR [ $r = 0.65$ ,  $p < 0.001$ , two-tailed]. Performance at 0 dB SNR also correlated positively with –3 dB SNR [ $r = 0.72$ ,  $p < 0.001$ , two-tailed] and –5 dB SNR [ $r = 0.48$ ,  $p = 0.016$ , two-tailed]. Finally, performance at –3 dB SNR correlated positively with –5 dB SNR [ $r = 0.45$ ,  $p = 0.023$ , two-tailed]. Thus, individual performance at one SNR was only related to that participant's performance on a similar SNR, and was not consistently related to performance at more or less favorable SNRs. This pattern suggests that meaningful individual differences in speech recognition abilities consistently emerged on PRESTO, compared with HINT, where performance varied more as a function of SNR. Mean word recognition scores on HINT at each SNR and a mean overall score across all SNRs are given in Table 1. Note again that because the overall distribution of trials by SNR differed for PRESTO and HINT, overall PRESTO and HINT scores should be compared with caution. Again, for the purpose of comparison, adjusted HINT scores were calculated from each listener's average of his or her four overall SNR scores, each of which contributed to 25% of the adjusted overall score. Adjusted scores were again similar to the overall scores across all trials for both non-native (mean adjusted overall HINT score = 39.0%) and native (mean adjusted overall HINT score = 60.9%) groups.

To compare non-native low-variability HINT sentence recognition with native HINT sentence recognition, we carried out a repeated-measures ANOVA on the word recognition scores with listener group (non-native, native) as the between-subject factor and SNR as a within-subject factor. The ANOVA revealed significant main effects for SNR [ $F_{(3,132)} = 547.26$ ,  $p < 0.001$ ] and listener group [ $F_{(1,44)} = 90.56$ ,  $p < 0.001$ ], as well as a significant SNR  $\times$  listener group interaction [ $F_{(3,132)} = 18.40$ ,  $p < 0.001$ ]. The significant main effect of SNR confirms that listeners were more accurate at more favorable SNRs, and will not be further analyzed. Figure 2 displays the average keyword accuracy on HINT by SNR for both the non-native and native listening groups. To explore the main effect of listener group and the SNR  $\times$  listener group interaction, we conducted independent-samples *t*-tests between groups for each SNR. The non-native listeners performed more poorly on HINT than the native listeners across all SNRs [ $t_{(44)} = 3.55$ – $9.36$ , all  $p < 0.001$ ]. Although non-native listeners were less accurate than the native listeners across all SNRs, unlike performance on the PRESTO sentences, the largest differences were not observed at the most favorable SNR (+3 dB SNR) but were observed at the moderate SNRs (0 and –3 dB SNR), as seen in Figure 2. Like PRESTO, however, the groups performed most similarly at –5 dB SNR.

**PRESTO versus HINT:** Comparing the performance on high-variability PRESTO sentences with low-variability HINT sentences, differences in performance on the tests vary

by SNR for both non-native listeners and native listeners. As reported in Gilbert et al (2013), for native listeners, PRESTO sentences were more difficult than HINT sentences at more favorable SNRs but were easier than HINT sentences at more difficult SNRs. This pattern is clearly visible in Figure 2, which displays PRESTO and HINT scores at each SNR. Whereas non-native performance on the PRESTO sentences was never greater than performance on HINT sentences, performance on the two sentence types was similar at the less favorable SNRs, suggesting that the benefit of understanding speech produced by a single talker compared with multiple talkers is limited at less favorable SNRs. Paired-comparison *t*-tests were carried out between PRESTO and HINT accuracy scores for each SNR. The analyses revealed that the non-native listeners performed more poorly on PRESTO across all SNRs, although the effect was only marginal at the less favorable SNRs, +3 dB SNR [ $t_{(24)} = 33.27$ ,  $p < 0.001$ ], 0 dB SNR [ $t_{(24)} = 8.34$ ,  $p < 0.001$ ], -3 dB SNR [ $t_{(24)} = 2.17$ ,  $p = 0.040$ ], and -5 dB SNR [ $t_{(24)} = 2.65$ ,  $p = 0.014$ ]. Although the PRESTO sentences never emerged as more intelligible for the non-native group, Figure 2 clearly shows that the differences between PRESTO and HINT were larger at more favorable SNRs. The mean difference at +3 dB SNR was 40.5% and 15.5% at 0 dB SNR, but the mean difference at -3 dB SNR was 3.9% and only 2.8% at -5 dB SNR, possibly reflecting floor effects for the non-native group.

Finally, on examination of group differences on PRESTO and HINT sentences, the difference between non-native and native performance was greater on PRESTO than on HINT. However, group differences also varied as a function of SNR. Table 1 shows the mean difference scores between the non-native and native groups on PRESTO and HINT. The largest differences between performance on PRESTO and HINT occurred at the most (+3 dB SNR) and least (-5 dB SNR) favorable presentation conditions. At +3 dB SNR, group differences were much larger on PRESTO than on HINT. At 0 and -3 dB SNR, differences were similar across tests, although PRESTO was slightly more challenging at -3 dB SNR (difference scores: 28.3% on PRESTO; 22.6% on HINT). At -5 dB SNR, group differences were again larger on PRESTO than on HINT. Across all SNRs, the difference scores show that non-native listeners were still consistently better on HINT than on PRESTO compared with the native listeners (difference scores averaged across all SNRs: 32.2% difference on PRESTO; and 21.8% on HINT).

### Indexical Processing Tasks

**Gender Discrimination:** Mean discrimination scores were calculated for the non-native and native listener groups for “Same-Gender” trials and “Different-Gender” trials, and overall. Mean accuracy on the Gender Discrimination task was very high, reflecting ceiling effects on this task. Overall accuracy for the non-native listeners was 96.8% (SD = 6.0%). Accuracy on the “Same-Gender” trials was 98.5% (SD = 4.1%), and accuracy on the “Different-Gender” trials was 95% (SD = 11.4%). Independent-samples *t*-tests were carried out between the non-native group and the native group on all three measures. No significant differences between listener groups were obtained on any of the gender discrimination measures. Mean RT measures were also calculated for correct trials. Independent-samples *t*-tests again revealed no significant differences between the non-native and native listeners on any of the RT measures. Thus, the non-native and native groups performed similarly in terms of both discrimination accuracy and RT on the Gender Discrimination task.

**Talker Discrimination:** Mean accuracy scores were calculated for non-native and native listener groups for “Same-Talker” trials and “Different-Talker” trials, and overall. Mean accuracy on the Talker Discrimination task was also very high, reflecting ceiling effects. Overall accuracy for the non-native listeners was 98.0% (SD = 3.2%). Accuracy on “Same-Talker” trials was 97.7% (SD = 4.5%), and accuracy on “Different-Gender” trials was 98.3% (SD = 3.4%). Independent-samples *t*-tests were carried out between the non-native and native listener groups. No significant differences were observed on any of the talker discrimination measures. Mean RT measures were also calculated for correct trials. Independent-samples *t*-tests again revealed no significant differences between the non-native and native listeners on any RT measure. As with gender discrimination, the non-native and native groups performed similarly in terms of response accuracy and RT on the Talker Discrimination task.

**Regional Dialect Categorization:** Mean categorization accuracy scores were calculated for the non-native and native listener groups for each of the six dialect regions and overall. Performance on the Regional Dialect Categorization task by the non-native listeners was very poor and close to chance (16.7%) for all six regional dialects as well as the overall score. Figure 3 shows the overall categorization accuracy and accuracy for each regional dialect by the two listener groups. A series of independent-sample *t*-tests between the non-native and native listener groups was carried out on overall accuracy and accuracy on each regional dialect. The *t*-tests revealed that the non-native group was significantly less accurate than the native group overall [ $t_{(44)} = 7.68, p < 0.001$ ] and on every regional dialect [ $t_{(44)} = 2.44\text{--}6.06, \text{all } p < 0.019$ ] except the Western dialect, where performance was poor for both listener groups.

**Self-Report Questionnaire on Vocabulary Knowledge—**Mean word familiarity ratings were calculated for both non-native and native listener groups for the low-, medium-, and high-familiarity words, and overall. As expected, a series of paired *t*-tests established that the non-native listeners gave higher ratings to the high-familiarity words than medium- [ $t_{(24)} = 10.61, p < 0.001$ ] and low-familiarity words [ $t_{(24)} = 9.89, p < 0.001$ ]. They also gave medium familiarity words higher ratings than they did for low-familiarity words [ $t_{(24)} = 4.42, p < 0.001$ ]. Independent-samples *t*-tests revealed that the non-native listeners gave lower ratings to the high-familiarity words [ $t_{(24)} = 4.35, p < 0.001$ ], and to all words overall on the WordFam questionnaire [ $t_{(24)} = 2.63, p = 0.012$ ] than did the native listeners. Figure 4 displays the average familiarity ratings for all three word types and overall for the two listener groups.

## Individual Differences: Correlation and Regression Analyses

**Demographic Factors—**Before examining individual differences in keyword recognition performance on the high-variability sentences, we assessed the influence of language background. The current study was designed to reduce the potential contributions of several background factors to speech recognition abilities by selecting only non-native listeners with the same native language and similar experience with the English language. As mentioned above, limiting the variability in language background allows us to more directly examine the effects of indexical processing skills and neurocognitive abilities on sentence



recognition. Given that the listeners shared similar language backgrounds, we expected that there would not be enough variability in their language backgrounds to significantly contribute to performance scores. To ensure that the traditional factors of length of residency, age of arrival in the U.S., and age of initial English instruction were not significantly related to individual L2 listeners' speech recognition abilities for this particular group of non-native listeners, we conducted several correlational analyses.

In our demographic assessment, participants provided some basic information about their experiences learning English and their residential histories in their home country and in the U.S. Length of residency ranged from approximately 1–27 mo. This demographic measure did not correlate significantly with any of the PRESTO measures. Age of initial learning of English ranged from 5–15 yr. Again, no significant relationships emerged between age of learning English and PRESTO performance. Age of arrival in the U.S. ranged from 18–33 yr. This measure correlated significantly with PRESTO performance at 0 dB SNR [ $r = 0.45$ ,  $p = 0.024$ , two-tailed], –3 dB SNR [ $r = 0.40$ ,  $p = 0.050$ , two-tailed], and overall [ $r = 0.41$ ,  $p = 0.044$ , two-tailed]. Listeners who were older when they arrived in the U.S. recognized more keywords on PRESTO, which was not expected. Given that all non-native listeners had arrived in the U.S. in their adulthood, and age of arrival was not related to the other demographic variables or vocabulary knowledge, it is likely that age of arrival simply reflects a proxy for variation in individual differences in students who move to the U.S. for schooling at an earlier age or a later age (e.g., undergraduate students or graduate students) and is not related to previous amount or type of experience and/or exposure to the English language. Therefore, age of arrival will not be analyzed further or considered in any of the analyses of individual differences on the indexical processing tasks, or ratings from the self-report questionnaires.

**Speech Recognition Tasks—**We conducted correlational analyses between keyword recognition accuracy on PRESTO and HINT sentence tests to examine the relationships between speech recognition in high-variability listening conditions and speech recognition in low-variability listening conditions. PRESTO keyword recognition scores at each SNR correlated with HINT performance at the same SNR [ $r = 0.48$ – $0.80$ , all  $p < 0.015$ , two-tailed]. Figure 5 shows the relationships between PRESTO and HINT scores for the four SNRs for the non-native listeners. Overall speech recognition performance on the high-variability PRESTO sentences also correlated strongly with speech recognition performance on the low-variability HINT sentences [ $r = 0.80$ ,  $p < 0.001$ ]. Thus, although overall non-native performance on PRESTO sentences compared with HINT varied greatly by SNR, as demonstrated in Figure 5, individual non-native listeners who were better at recognizing highly variable speech in multitalker babble were also consistently better at recognizing less variable speech in multitalker babble.

**Sources of Individual Differences on PRESTO—**To explore the factors contributing to individual differences in speech recognition in high-variability listening conditions, we carried out a series of stepwise multiple linear regression analyses to identify some possible contributors to L2 high-variability speech recognition abilities. Taking the results of the group analyses into consideration, we used the following measures as predictor variables in

the analyses: Regional Dialect Categorization accuracy, Gender Discrimination accuracy and RT, Talker Discrimination accuracy and RT, WordFam overall mean word familiarity ratings, and the overall global GEC score from the BRIEF-A questionnaire. Finally, length of residency and age of initial learning of English were also included as measures of L2 proficiency.

Given that sentence recognition accuracy across all SNRs correlated highly, the overall mean accuracy score across all conditions was first analyzed. Only the global GEC executive functioning score for the BRIEF-A emerged as a significant predictor [ $R^2 = 0.24$ ,  $F_{(1, 23)} = 7.43$ ,  $p = 0.012$ ]. None of the other perception or demographic variables was significant. Similar analyses were also carried out for PRESTO keyword recognition accuracy for each SNR (+3 dB, 0 dB, -3 dB, -5 dB SNR) to assess whether different sets of skills are related to understanding the variable speech in different SNR conditions. At +3 dB SNR, GEC again emerged as the only significant predictor [ $R^2 = 0.22$ ,  $F_{(1, 23)} = 6.46$ ,  $p = 0.018$ ]. Similarly, at 0 dB SNR, GEC was again the only significant predictor [ $R^2 = 0.36$ ,  $F_{(1, 23)} = 12.73$ ,  $p = 0.002$ ]. No factor emerged as significant for -3 and -5 dB SNR, where performance was low and less variable among individual listeners.

The results of the multiple regression analyses suggest that executive functioning may play an important role in recognizing speech in the PRESTO sentences, especially at more moderate SNRs. To further explore the relationship between executive functioning and high-variability speech recognition, we performed a series of bivariate correlations between PRESTO keyword accuracy scores at +3 and 0 dB SNR and BRIEF-A measures. Correlational analyses revealed numerous significant correlations between keyword recognition scores and self-report measures of executive functioning on the BRIEF-A. Performance on PRESTO across all SNRs correlated significantly with almost every BRIEF-A clinical subscale as well as the three aggregate composite scores. Table 2 displays a summary of the correlational analyses between percent-correct keyword recognition on PRESTO scores (overall, and at +3 and 0 dB SNR) and BRIEF-A measures. All reported correlations are one-tailed, given the predication that participants with better executive functioning abilities would be better able to rapidly adapt and adjust to variability in the PRESTO sentences. Also note that all correlations were negative because higher scores on the BRIEF-A reflect greater disturbances in a particular executive function domain. Although multiple tests were carried out, because these analyses are exploratory in nature, results of all tests—including those that emerged as non-significant or only marginally significant—are reported. As shown in Table 2, the relationship between executive functioning and PRESTO scores is generally consistent across all BRIEF-A subscales, indicating the pervasive contribution of executive function to the recognition of keywords in high-variability sentences for non-native listeners.

## DISCUSSION

The current study was designed to investigate non-native speech recognition in high-variability listening conditions and to identify several underlying information processing factors that are associated with individual differences in speech recognition abilities. To assess the effects of high-variability sentence materials on non-native speech recognition, 25

native speakers of Mandarin living in Bloomington, IN, completed two sentence recognition tasks using sentences from a novel high-variability sentence recognition test (PRESTO) and sentences from a conventional single-talker low-variability sentence test (HINT). Despite being highly educated undergraduate and graduate students, non-native listeners tested in this study displayed a great deal of difficulty recognizing PRESTO sentences in multitalker babble. Mean percent-correct keyword recognition accuracy on PRESTO across all SNRs was only 23.2% ( $SD = 5.7\%$ ). As expected, performance was also worse at less favorable SNRs. Compared with PRESTO sentences in multitalker babble, non-native listeners performed better on the single-talker HINT sentences in multitalker babble. Across all SNRs, non-native listeners recognized more words in HINT sentences than PRESTO sentences, but the differences varied as a function of SNR with non-native listeners performing much better on HINT at more favorable SNRs than at less favorable SNRs. This finding suggests that the additional indexical variability from different talkers in the PRESTO sentences consistently creates more challenging listening conditions for non-native listeners. Although non-native listeners benefitted from the presentation of the same talker on HINT at more favorable SNRs, allowing them to learn and use talker-contingent information encoded in the signal to benefit speech recognition, the non-native listeners were unable to rapidly adjust and optimally adapt to the highly variable changing test sentences on PRESTO from trial to trial. These results provide further support for the results reported in earlier studies showing that listeners benefit substantially from talker-contingent perceptual learning in speech recognition (e.g., Nygaard et al, 1994; Nygaard and Pisoni, 1998), as well as other reports showing that non-native listeners have much more difficulty understanding speech produced by talkers with different regional dialects or foreign accents (e.g., Eisenstein and Verdi, 1985; Fox and McGory, 2007; Ikeno and Hansen, 2007; Cooke et al, 2008).

A comparison of non-native listeners' keyword recognition accuracy on PRESTO and HINT with native listeners' scores, however, revealed that the relative difficulty of high-variability PRESTO sentences for the non-native listeners differed from the native listeners as a function of SNR. Difference scores on PRESTO and HINT across all SNRs showed that the non-native listeners were consistently better on HINT than on PRESTO compared with native listeners, suggesting that high-variability sentences were disproportionately more difficult for the non-native listeners compared with the native listeners, even for highly educated, proficient non-native speakers of English. The disproportionate effect of high-variability sentences for non-native listeners was expected, because the non-native listeners have less exposure to and experience with sociolinguistic variability in the L2 as well as less knowledge of L2 grammatical structures and words.

However, the difference scores also showed an unexpected interaction of sentence type (PRESTO and HINT) and SNR. Previous findings reported in the literature have suggested that performance (for both native and non-native listeners) becomes worse in conditions that are more challenging (e.g., less favorable SNRs, unfamiliar accents or dialects, multiple talkers, etc.), and that non-native listeners' performance becomes disproportionately worse compared with that of native listeners (e.g., Florentine et al, 1984; Eisenstein and Verdi, 1985; Florentine, 1985; Buus et al, 1986; van Wijngaarden et al, 2002). Given these previous studies, it was expected that PRESTO difference scores would be greater than

HINT difference scores at all SNRs, and that difference scores at less favorable SNRs would be greater than difference scores at more favorable SNRs. However, in the current study, PRESTO was disproportionately more difficult than HINT for non-native listeners (i.e., PRESTO difference scores were greater than HINT difference scores) at the most and least favorable SNRs, not just at the least favorable conditions as would be predicted based on earlier findings. In addition, non-native and native performance on PRESTO was more similar among groups at less favorable SNRs (i.e., PRESTO difference scores smaller), and performance on HINT was most similar at most (+3 dB) and least (−5 dB) favorable SNRs, not just at the less favorable SNRs as would be predicted.

Two factors may be responsible for these patterns. First, the small difference scores at the more challenging SNRs may simply reflect floor effects. Second, the observed interaction showing that difference scores were greater for PRESTO than for HINT at the most favorable SNR may be related to the differential task demands resulting from the increased indexical variability on PRESTO. For HINT sentences at more favorable SNRs, the non-native listeners likely benefitted from the lack of talker variability, making efficient use of talker-specific and context information to recognize the words in these sentences, but not at the poorer SNRs, where the talker-specific information may not have been audible. On PRESTO, however, although the more moderate SNR allows the listener to recognize some additional keywords, the variability in the signal and vocal source from trial to trial attenuates the learning of talker-specific information and the use of context information to benefit speech recognition (Gilbert et al, 2013). Thus, although the details of the findings obtained with the non-native listeners are inconsistent with previous studies examining speech recognition in adverse listening conditions, the results of the present study can be accounted for by considering the combined influence of both the variability in the speech signals used in PRESTO and the background competition from multitalker babble.

The non-native listeners also completed several additional perceptual tasks that were designed to investigate the processes underlying non-native perception of indexical variability in speech. The non-native participants performed close to ceiling and were similar to the native control groups on both the Talker Discrimination and Gender Discrimination tasks. Thus, the poor performance observed on PRESTO by the non-native group is not related in any simple way to audibility of the signals or a listener's ability to detect and discriminate indexical information associated with talker or gender differences in spoken sentences presented in quiet. The non-native listeners in this study did not have any difficulty discriminating gender-specific or talker-specific differences in English sentences. Additionally, we failed to find any significant correlations between PRESTO accuracy and Gender Discrimination and Talker Discrimination accuracy or RTs because of ceiling effects and the absence of variance in these scores.

These findings are consistent with those of Cooke et al (2008), who found that non-native listeners were equally sensitive to gender-specific information as native listeners, suggesting that low speech recognition accuracy on PRESTO for the non-native listeners is not related to audibility or early sensory or perceptual encoding problems but, rather, is related to incomplete English phonological systems and lack of lexical knowledge of English words. Although some non-native listeners, like native listeners, may also have difficulty encoding

indexical variability in either their native language or their L2, lack of experience with and exposure to the L2 result in much greater difficulties in spoken word recognition in an L2 in high-variability conditions using PRESTO sentences.

The results of the Regional Dialect Categorization task also revealed that a lack of experience with different varieties of American English is particularly harmful to the encoding and processing of indexical information in English. The listeners in the current study were very poor at categorizing the talkers by region of origin. Mean accuracy in the six-alternative forced-choice dialect categorization task was at chance. Clopper and Bradlow (2009) reported that non-native listeners were able to perceive and discriminate dialect-specific acoustic-phonetic information in their L2 in an auditory free classification task. Although the results of the two studies may seem contradictory, the differences may simply reflect differential task demands of the two procedures. The free classification task permits listeners to use acoustic-phonetic information present in the target sentences to form their own unique perceptual categories based on regional dialect perceptual similarity. The more conventional forced-choice categorization task, on the other hand, does not allow listeners to create their own perceptual categories or to repeatedly compare different talkers to each other. Instead, the listener must use the predefined response categories provided by the experimenter, which rely on stored knowledge of dialect-specific differences to categorize a talker after only a single exposure to the target sentence. Because of these important task differences, the poor performance of the non-native listeners on the forced-choice categorization task in the present study is likely the result of lack of experience and knowledge of the different regional dialects of American English and the absence of robust stable dialect categories in long-term memory and does not reflect a sensory processing deficit related to the perception (detection and discrimination) of indexical information in the speech signal. However, individual variability in the ability to perceive and encode dialect-specific information may be reflected in performance differences in the non-native group. The relationship between the task demands of the free classification task and forced-choice regional dialect categorization task should be further examined to uncover the factors that contribute to differential performance, and to understand more precisely how regional dialect processing skills are related to sentence recognition performance in noise.

Finally, the results of the analysis of word familiarity scores in English suggest that differences in lexical knowledge and lexical connectivity of words in long-term memory may also contribute to speech recognition difficulties of non-native listeners. Non-native listeners had lower familiarity scores on the high-familiarity words than did the native listeners in our earlier study. This result may reflect the types of words non-native students at a university in the U.S. learn, or have been taught before their arrival in the U.S. Because our participants were all highly educated undergraduate and graduate students, it is likely that they were exposed to many difficult and unfamiliar English words. However, although the non-native listeners had been exposed to the low- and medium-familiarity words, the significant difference between the non-native and native groups for the high-familiarity words suggests a lack of native-like exposure to common English words and overall less lexical connectivity in long-term memory. Thus, non-native listeners' overall less familiarity with English may have contributed to the poorer performance observed for the non-native listeners on PRESTO. Differences in WordFam scores obtained here replicate earlier studies

demonstrating the important contribution of lexical knowledge in speech perception in an L2 (e.g., Bradlow and Pisoni, 1999; Ezzatian et al, 2010; Bundgaard-Nielsen et al, 2011), and recognizing highly variable speech (Tamati et al, 2013). Taken together, findings from the current study and previous studies suggest that poor speech recognition abilities in high-variability listening conditions in noise may be associated with the use of higher-order context based on lexical knowledge and the organization of words in lexical memory (e.g., Pisoni et al, 1985; Altieri et al, 2010). The use of lexical and sublexical knowledge and lexical connectivity in long-term memory should be further explored using other types of behavioral measures to assess vocabulary knowledge and links among words in lexical memory in both native and non-native listeners (e.g., Beckage et al, 2011).

Individual differences in the non-native group were also investigated. In particular, indexical processing, vocabulary knowledge, neurocognitive abilities, and language background measures were obtained from the non-native listeners. Given that variability in talker, gender, and regional dialect are core components of PRESTO sentences, we also expected that listeners who are better at recognizing the high-variability PRESTO sentences would also be better on other indexical processing tasks that directly assess these component processes. However, none of the indexical processing measures emerged as significant predictors of high-variability PRESTO keyword recognition scores, which may be the result of ceiling effects for the talker and gender discrimination and floor effects for regional dialect categorization. Individual differences in the perception of indexical information in speech in the L2 should be investigated in greater depth in future studies using different converging methods and L2 listener groups with different language backgrounds, including those who may have more exposure to regional dialectal variability in their L2 because of longer lengths of residency or greater interaction with native speakers.

Vocabulary knowledge, as measured by the WordFam self-report questionnaire, also was not found to be related to PRESTO performance. This result was surprising given the earlier results of Tamati et al (2013), and other research findings demonstrating the influence of lexical knowledge in an L2 (e.g., Bradlow and Pisoni, 1999; Ezzatian et al, 2010; Bundgaard-Nielsen et al, 2011). It is possible that vocabulary knowledge did not emerge as a significant predictor of high-variability keyword recognition accuracy because the WordFam scores for the non-native listeners were very low and displayed little individual variability. The non-native listeners either know the same set of words, or the words on which they differ were not assessed by this questionnaire, which was developed and normed for use by native English listeners. Another possibility is that knowing more words does not help the recognition process when the sentence recognition task is very difficult, as it was for the non-native listeners in this study. The adverse conditions incorporated in the PRESTO sentence materials were so difficult for the non-native listeners that they may not be able to benefit from top-down lexical knowledge to support robust speech recognition. The contribution of lexicon size and lexical connectivity to individual differences in speech recognition in adverse listening conditions should be further explored in other groups of non-native listeners who have more exposure and more diverse experiences with English to assess the effects of L2 experience on sentence recognition in noise and multitalker babble.



The non-native listeners also completed the BRIEF-A behavioral rating questionnaire that was used to assess executive functioning and behavioral regulation skills in everyday, real-world situations. Given previous research on the relationship between neurocognitive processes and speech perception skills (e.g., Pisoni, 2000; Arlinger et al, 2009; Stenfelt and Rönnberg, 2009), we expected that the non-native listeners who were better at recognizing PRESTO sentences would also show better executive functioning skills. The global GEC executive functioning score emerged as significant in the analyses of individual differences in performance on the PRESTO sentences. Follow-up correlational analyses also revealed numerous significant correlations between PRESTO keyword accuracy and the BRIEF-A clinical subscales with the non-native listeners. In our previous study, we suggested that the BRIEF-A questionnaire may not be ideal for identifying individual differences in executive functioning in highly educated, healthy young adults because it was designed to be used with individuals from an attention- deficit/hyperactivity disorder clinical population who are having significant difficulties in everyday situations and is not a general-purpose instrument for use with individuals within the normal limits of executive functioning abilities (Tamati et al, 2013). However, the BRIEF-A was, in fact, sensitive to individual differences in this group of non-native listeners and the BRIEF-A scores revealed several important correlations, suggesting that executive function and cognitive control processes contribute to speech recognition performance in high-variability listening conditions in this population.

We found that self-reports of executive functioning were related to PRESTO scores at the more favorable SNRs tested, where recognition was relatively easier for the non-native listener group. The results obtained with the BRIEF-A suggest that the ability to rapidly adjust to continuously changing test sentences against background competition from multitalker babble is related to individual differences in several foundational component executive functioning skills—specifically in domains such as working memory, inhibition, and shifting—and replicates findings from previous studies demonstrating close links between basic core neurocognitive abilities and speech perception skills in children (e.g., Pisoni and Geers, 2000; Cleary et al, 2000; Pisoni and Cleary, 2003; Beer et al, 2011; Pisoni et al, 2011) and adults (e.g., van Rooij and Plomp, 1990; Gordon-Salant and Fitzgibbons, 1997; Humes, 2007; Akeroyd, 2008; Conway et al, 2010; Jesse and Janse, 2012; Tamati et al, 2013).

Recently, Darcy et al (2012) also reported significant correlations between BRIEF-A measures (from a version of the BRIEF-A translated into their participants' native language) and performance on L2 phonological processing tasks. One reason why the BRIEF-A questionnaire may be more useful and informative with non-native speakers, despite being created to measure difficulties related to executive functioning of individuals with attention-deficit/hyperactivity disorder (Roth et al, 2005), is the stress and daily demands of living in a foreign country and trying to cope and manage everyday tasks in another language and in a different language environment. The challenges of everyday life and activities of daily living for a non-native speaker in a novel foreign country may accentuate existing underlying difficulties in executive functioning, cognitive control, and behavior regulation, making the individual more self-aware of problems or difficulties. Additionally, because we have a diverse group of international students in both undergraduate and graduate programs, as mentioned above, individuals in the non-native participant group may have a large range of

L2 grammatical and lexical knowledge, as well as cognitive abilities, including executive functioning skills. The influence of experience and exposure in the L2 environment on individuals' executive functioning skills, and their perception of their own executive functioning, should be further investigated with non-native listeners with a wider range of length of residency and L2 proficiency. Additionally, future studies of non-native speech recognition in adverse listening conditions should include performance-based tests to assess executive functioning and cognitive control processes to further evaluate the contribution of executive function and cognitive control processes in speech perception in high-variability conditions.

Because language background and prior linguistic experience have been shown to contribute to speech perception abilities in individual L2 learners, one goal of the current study was to limit the potential number of contributing factors from language background to focus more on individual differences in speech recognition in highly variable conditions. Therefore, only native speakers of Mandarin, from mainland China, with short lengths of residencies in a limited number of locations in the U.S. were selected for participation. Correlational analyses failed to reveal any significant relationships between length of residency or age when the non-native listeners started learning English and keyword recognition on PRESTO. However, the listeners' age of arrival in the U.S. was found to correlate significantly with PRESTO performance. The older a participant was when he or she arrived in the U.S., the better he or she performed on PRESTO sentences. Although this finding seems at first counterintuitive, one explanation of this result is that all of the participants were students at Indiana University at the time of testing, and the listeners who had arrived later were generally older students attending graduate school. Because age of arrival was not related to the other demographic variables or vocabulary knowledge, it may reflect individual differences in attitudes, behavioral traits, and core neurocognitive abilities, rather than previous linguistic experience or knowledge. Nevertheless, factoring out the variability in exposure to English allows us to take the complex relationships between the demographic variables and performance measures into consideration and still be able to interpret the individual influences of indexical processing skills and neurocognitive abilities on high-variability speech recognition performance. Indeed, these demographic variables may play a larger role in speech recognition in high-variability conditions with a more diverse group of non-native listeners, and these effects are simply limited in the current study. Attempts should be made in future studies to assess the contribution of these demographic variables to speech recognition in high-variability conditions, including individual listeners and groups with more diverse language backgrounds.

Taken together, the results of the current study suggest that natural variability in speech contributes to more difficult listening conditions for non-native listeners. When considering the difficulties of non-native listeners in L2 communicative settings, the types and amounts of test variability, beyond audibility or energetic and/or informational masking, should also be considered. The current results suggest that lack of explicit knowledge of regional dialect variability in the U.S. and less knowledge of English vocabulary contribute to the non-native listeners' difficulty with the high-variability PRESTO materials. Other factors outside the scope of the current study (e.g., L2 phonological knowledge) also warrant further investigation. Identifying sources of variability in speech as a potential challenge to non-

native speech recognition can help direct targeted instruction and can serve as a basis to motivate the development of novel training programs to improve speech recognition and understanding in non-native listeners.

In the current study, non-native speech perception performance in high-variability conditions also revealed substantial amounts of individual variability. The contribution of factors such as experience with L2 phonology and/or exposure to multiple sources of indexical variability in an L2 may be critical to explaining observed differences, but individual differences in neurocognitive abilities in domains, such as executive functioning and cognitive control processes, may also contribute to robust spoken-word recognition abilities in adverse listening conditions. The current study used a limited test battery covering a narrow range of information-processing skills to begin to elucidate the many potential sources of individual differences. Variability in non-native speech recognition skills likely reflects the use of multiple neurocognitive processes and associated perceptual and cognitive abilities working together as an integrated system. Indexical processing skills of non-native listeners using other sources of indexical information, such as a foreign accent or emotions in speech, as well as tests of L2 phonological awareness should be investigated using a broad range of information processing and psycholinguistic tasks in various settings or conditions. Similarly, other neurocognitive performance measures, such as cognitive control processes used in attention/inhibition, short-term and working memory, and processing speed, should be assessed in future studies of individual differences in non-native speech recognition.

## CONCLUSIONS

The goal of the current study was to investigate non-native speech recognition in high-variability listening conditions. The results revealed that high-variability conditions were particularly challenging for non-native listeners, especially at less favorable SNRs, compared with a group of native listeners who were tested in the same listening conditions. Thus, variability in speech leads to more challenging listening conditions, not only for native listeners, but also, and perhaps even more so, for non-native listeners who have incomplete knowledge and access to phonological, lexical, and syntactic information of the L2. Findings from the indexical processing tasks demonstrated that although non-native listeners were able to easily discriminate acoustic-phonetic information related to talker or gender differences, they lacked explicit detailed knowledge of regional dialects in the U.S. to reliably categorize talkers by their residential history. Performance on these indexical processing tasks, taken together with differences in vocabulary knowledge between the non-native and native listeners, suggests that difficulties in processing high-variability speech is related to lack of knowledge and exposure to variability in an L2. Investigation of individual differences among the non-native listeners also revealed that executive functioning abilities measured by the BRIEF-A self-report inventory contributed to observed individual differences in speech recognition abilities, particularly among non-native listeners living in the U.S. The results of the current investigation suggest several new and promising directions for future research to identify potential are as for intervention and focused training within and outside the L2 classroom. Additional research on the perception of indexical variability in an L2, and the neurocognitive factors underlying individual differences in L2

speech perception abilities, should be carried out using additional experimental methods and other populations of non-native listeners to replicate and extend the present set of findings.

## Acknowledgments

Preparation of this manuscript was supported in part by National Institutes of Health National Institute on Deafness and Other Communication Disorders (NIH NIDCD) Training Grant T32DC00012 and NIH NIDCD Research Grant R01-DC00111 to Indiana University.

The authors thank Jaimie Gilbert, Luis Hernandez, Katie Faulkner, Emily Garl, Lindsay Stone, and Marissa Habeshy for their assistance on this project.

## Abbreviations

<b>BRI</b>	Behavioral Regulation Index
<b>BRIEF-A</b>	Behavioral Rating Inventory of Executive Function – Adult Version
<b>GEC</b>	General Executive Composite score from the BRIEF-A
<b>HINT</b>	Hearing In Noise Test
<b>L2</b>	second language
<b>MI</b>	Metacognitive Index
<b>PRESTO</b>	Perceptually Robust English Sentence Test Open-set
<b>RT</b>	response time
<b>SNR</b>	signal-to-noise ratio
<b>U.S</b>	United States

## References

- Abercrombie, D. Elements of General Phonetics. Chicago, IL: Aldine; 1967.
- Akeroyd MA. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol*. 2008; 47(Suppl 2):S53–S71. [PubMed: 19012113]
- Altieri N, Gruenenfelder T, Pisoni DB. Clustering coefficients of lexical neighborhoods: Does neighborhood structure matter in spoken word recognition? *Ment Lex*. 2010; 5(1):1–21. [PubMed: 21423865]
- Arlinger S, Lunner T, Lyxell B, Pichora-Fuller MK. The emergence of cognitive hearing science. *Scand J Psychol*. 2009; 50(5):371–384. [PubMed: 19778385]
- Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol Bull*. 1997; 121(1):65–94. [PubMed: 9000892]
- Barkley, RA. Executive Functions: What They Are, How They Work, and Why They Evolved. New York: The Guilford Press; 2012.
- Beer J, Kronenberger WG, Pisoni DB. Executive function in everyday life: implications for young cochlear implant users. *Cochlear Implants Int*. 2011; 12(Suppl 1):S89–S91. [PubMed: 21756484]
- Beckage N, Smith L, Hills T. Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*. 2011; 6(5):e19348. [PubMed: 21589924]
- Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol*. 1979; 13(3):108–112. [PubMed: 486816]
- Bent T, Bradlow AR. The inter language speech intelligibility benefit. *J Acoust Soc Am*. 2003; 114(3):1600–1610. [PubMed: 14514213]

- Black JW, Hast MH. Speech reception with altering signal. *J Speech Hear Res.* 1962; 5:70–75. [PubMed: 13869656]
- Bradlow AR, Pisoni DB. Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors. *J Acoust Soc Am.* 1999; 106(4 Pt 1):2074–2085. [PubMed: 10530030]
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun.* 1996; 20(3):255–272. [PubMed: 21461127]
- Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am.* 2001; 109(3):1101–1109. [PubMed: 11303924]
- Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am.* 2001; 110(5 Pt 1):2527–2538. [PubMed: 11757942]
- Bundgaard-Nielsen RL, Best CT, Tyler MD. Vocabulary size matters: the assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Appl Psycholinguist.* 2011; 32:51–67.
- Buus, S.; Florentine, M.; Scharf, B.; Canevet, G. Native French listeners' perception of American-English in noise. *Proceedings of Inter Noise 86*; Cambridge, MA. 1986. p. 895-898.
- Calandruccio L, Dhar S, Bradlow AR. Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J Acoust Soc Am.* 2010; 128(2):860–869. [PubMed: 20707455]
- Carhart R, Johnson C, Goodman J. Perceptual masking of spondees by combinations of talkers. *J Acoust Soc Am.* 1975; 58 (Suppl 1):S35.
- Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am.* 1969; 45(3):694–703. [PubMed: 5776931]
- Cleary M, Pisoni DB. Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *Ann Otol Rhinol Laryngol Suppl.* 2002; 189:113–118. [PubMed: 12018337]
- Cleary M, Pisoni DB, Kirk KI. Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. *Volta Review.* 2000; 102(4):259–280. [PubMed: 21666765]
- Cleary M, Pisoni DB, Kirk KI. Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *J Speech Lang Hear Res.* 2005; 48(1):204–223. [PubMed: 15938065]
- Clopper CG, Bradlow AR. Perception of dialect variation in noise: intelligibility and classification. *Lang Speech.* 2008; 51(Pt 3):175–198. [PubMed: 19626923]
- Clopper CG, Bradlow AR. Free classification of American English dialects by native and non-native listeners. *J Phonetics.* 2009; 37(4):436–451.
- Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *J Phonetics.* 2004b; 32(1):111–140.
- Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: word predictability is the key. *Cognition.* 2010; 114(3):356–371. [PubMed: 19922909]
- Cooke M, García Lecumberri ML, Barker J. The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am.* 2008; 123 (1):414–427. [PubMed: 18177170]
- Creelman CD. Case of the unknown talker. *J Acoust Soc Am.* 1957; 67:253–361.
- Cutler A, Weber A, Smits R, Cooper N. Patterns of English phoneme confusions by native and non-native listeners. *J Acoust Soc Am.* 2004; 116(6):3668–3678. [PubMed: 15658717]
- Darcy, I.; Park, H-Y.; Yang, C-L.; Gleiser, A. Individual differences in the development of L2 phonological processing: The contribution of cognitive abilities and executive function. Poster presented at the Conference on Sources of Individual Linguistic Differences; Ottawa, Canada. Mar. 2012
- Eisenstein M, Verdi G. The intelligibility of social dialects for working-class adult learners of English. *Lang Learn.* 1985; 35(2):287–298.

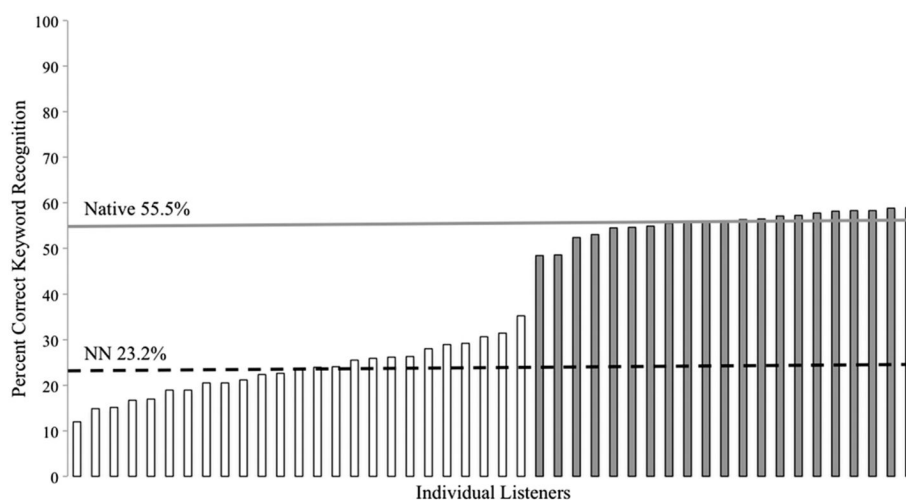
- Ezzatian P, Avivi M, Schneider BA. Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? *Speech Commun.* 2010; 52(11–12):919–929.
- Elliott LL. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *J Acoust Soc Am.* 1979; 66(3):651–653. [PubMed: 489836]
- Fallon M, Trehub SE, Schneider BA. Children's perception of speech in multitalker babble. *J Acoust Soc Am.* 2000; 108(6):3023–3029. [PubMed: 11144594]
- Felty, R. Unpublished manuscript. Indiana University; Bloomington: 2008. Perceptually Robust English Sentence Test (Open-Set).
- Finitzo-Hieber T, Tillman TW. Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *J Speech Hear Res.* 1978; 21(3):440–458. [PubMed: 713515]
- Flege JE, Bohn O-S, Jang S. Effects of experience on non-native speakers' production and perception of English vowels. *J Phonetics.* 1997; 25(4):437–470.
- Flege JE, Liu S. The effect of experience on adults' acquisition of a second language. *Stud Sec Lang Acquis.* 2001; 23:527–552.
- Flege JE, MacKay IRA. Perceiving vowels in a second language. *Stud Sec Lang Acquis.* 2004; 26(1): 1–34.
- Florentine, M. Non-native listeners' perception of American-English in noise. *Proceedings of Inter Noise 85*; Munich, Germany. 1985. p. 1021–1024.
- Florentine M, Buus S, Scharf B, Canevet G. Speech reception thresholds in noise for native and non-native listeners. *J Acoust Soc Am.* 1984; 75:S84.
- Fox, RA.; McGory, JT. Second language acquisition of a regional dialect of American English by native Japanese speakers. In: Bohn, O-S.; Munro, MJ., editors. *Language Experience in Second Language Speech Learning*. Amsterdam, the Netherlands: John Benjamins; 2007. p. 117–143.
- Ganong WF 3rd. Phonetic categorization in auditory word perception. *J Exp Psychol Hum Percept Perform.* 1980; 6(1):110–125. [PubMed: 6444985]
- García Lecumberri ML, Cooke M. Effect of masker type on native and non-native consonant perception in noise. *J Acoust Soc Am.* 2006; 119(4):2445–2454. [PubMed: 16642857]
- García Lecumberri ML, Cooke M, Cutler A. Non-native speech perception in adverse conditions: A review. *Speech Commun.* 2010; 52:864–886.
- Garofolo, JS.; Lamel, LF.; Fisher, WM.; Fiscus, JG.; Pallett, DS.; Dahlgren, NL. *The DARPA TIMIT acoustic-phonetic continuous speech corpus*. Philadelphia, PA: Linguistic Data Consortium; 1993.
- Gat IB, Keith RW. An effect of linguistic experience. Auditory word discrimination by native and non-native speakers of English. *Audiology.* 1978; 17(4):339–345. [PubMed: 687239]
- Gilbert JL, Tamati TN, Pisoni DB. Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *J Am Acad Audiol.* 2013; 24(1):26–36. [PubMed: 23231814]
- Gordon-Salant S, Fitzgibbons PJ. Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index. *J Speech Hear Res.* 1995a; 38(3):706–713. [PubMed: 7674661]
- Gordon-Salant S, Fitzgibbons PJ. Recognition of multiply degraded speech by young and elderly listeners. *J Speech Hear Res.* 1995b; 38(5):1150–1156. [PubMed: 8558883]
- Gordon-Salant S, Fitzgibbons PJ. Selected cognitive factors and speech recognition performance among young and elderly listeners. *J Speech Lang Hear Res.* 1997; 40(2):423–431. [PubMed: 9130210]
- Helfer KS, Freyman RL. Lexical and indexical cues in masking by competing speech. *J Acoust Soc Am.* 2009; 125(1):447–456. [PubMed: 19173430]
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007; 18(7):590–603. [PubMed: 18236646]
- Humes LE, Roberts L. Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility. *J Speech Hear Res.* 1990; 33(4):726–735. [PubMed: 2273886]
- Ikeno A, Hansen JHL. The effect of listener accent background on accent perception and comprehension. *EURASIP J Au Sp Mus Proc.* 2007; 2007:076030.



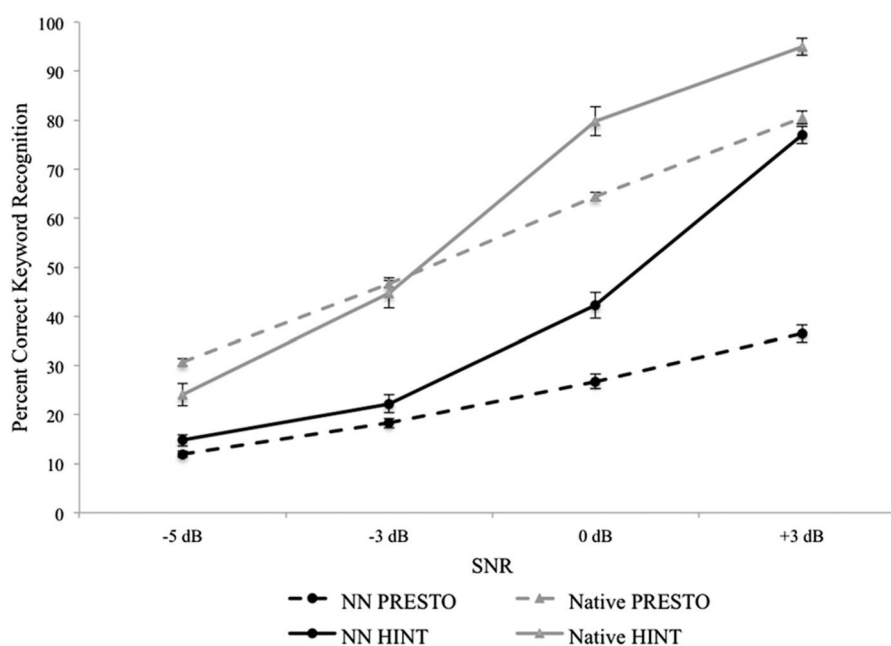
- Imai S, Walley AC, Flege JE. Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *J Acoust Soc Am*. 2005; 117(2):896–907. [PubMed: 15759709]
- Ingvalson EM, McClelland JL, Holt LL. Predicting native English-like performance by native Japanese speakers. *J Phonetics*. 2011; 39(4):571–584.
- Jenkins, JJ. Four points to remember: A tetrahedral model of memory experiments. In: Cermak, LS.; Craik, FIM., editors. *Levels of Processing in Human Memory*. Hillsdale, NJ: Erlbaum; 1979. p. 429-446.
- Jesse A, Janse E. Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Lang Cogn Process*. 2012; 27(7–8):1167–1191.
- Johnson, K.; Mullennix, JW. Complex representations used in speech processing: Overview of the book. In: Johnson, K.; Mullennix, JW., editors. *Talker Variability in Speech Processing*. San Diego, CA: Academic Press; 1997. p. 1-8.
- Karl JR, Pisoni DB. The role of talker-specific information in memory for spoken sentences. *J Acoust Soc Am*. 1994; 95:2873.
- Kreiman, J.; Van Lancker Sidsit, D. *Foundations of Voice Studies: An Inter disciplinary Approach to Voice Production and Perception*. Boston, MA: Wiley-Blackwell; 2011.
- Labov, W.; Ash, S. Understanding Birmingham. In: Bernstein, T.; Nunnally, T.; Sabino, R., editors. *Language Variety in the South Revisited*. Tuscaloosa, AL: Alabama University Press; 1997. p. 508-573.
- Lane H. Foreign accent and speech distortion. *J Acoust Soc Am*. 1963; 35:451–453.
- Lass NJ, Hughes KR, Bowyer MD, Waters LT, Bourne VT. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J Acoust Soc Am*. 1976; 59(3):675–678. [PubMed: 1254794]
- Lewellen MJ, Goldinger SD, Pisoni DB, Greene BG. Lexical familiarity and processing efficiency: individual differences in naming, lexical decision, and semantic categorization. *J Exp Psychol Gen*. 1993; 122(3):316–330. [PubMed: 8371087]
- MacKay IRA, Flege JE, Piske T, Schirru C. Category restructuring during second-language speech acquisition. *J Acoust Soc Am*. 2001a; 110(1):516–528. [PubMed: 11508976]
- MacKay IRA, Meador D, Flege JE. The identification of English consonants by native speakers of Italian. *Phonetica*. 2001b; 58 (1–2):103–125. [PubMed: 11096371]
- Mason HM. Understand ability of speech in noise as affected by region of origin of speaker and listener. *Speech Monographs*. 1946; 13 (2):54–58.
- Mattys SL, Davis MH, Bradlow AR, Scott SK. Speech recognition in adverse conditions: A review. *Lang Cogn Process*. 2012; 27(7–8):953–978.
- Mayo LH, Florentine M, Buus S. Age of second-language acquisition and perception of speech in noise. *J Speech Lang Hear Res*. 1997; 40(3):686–693. [PubMed: 9210123]
- McGarr NS. The intelligibility of deaf speech to experienced and inexperienced listeners. *J Speech Hear Res*. 1983; 26(3):451–458. [PubMed: 6645470]
- Meador D, Flege JE, MacKay IRA. Factors affecting the recognition of words in a second language. *Biling Lang Cogn*. 2000; 3(1):55–67.
- Miyake, A.; Friedman, NP. Individual differences in second language proficiency: working memory as language aptitude. In: Healy, AF.; Bourne, LE., Jr, editors. *Foreign Language learning: Psycholinguistic studies on Training and Retention*. Mahawah, NJ: Lawrence Erlbaum; 1998. p. 339-364.
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *J Acoust Soc Am*. 1989; 85(1):365–378. [PubMed: 2921419]
- Nábělek AK, Donahue AM. Perception of consonants in reverberation by native and non-native listeners. *J Acoust Soc Am*. 1984; 75(2):632–634. [PubMed: 6699302]
- Neff DL, Dethlefs TM. Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *J Acoust Soc Am*. 1995; 98(1):125–134. [PubMed: 7608391]
- Neuman AC, Hochberg I. Children’s perception of speech in reverberation. *J Acoust Soc Am*. 1983; 73(6):2145–2149. [PubMed: 6875100]

- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*. 1994; 95(2):1085–1099. [PubMed: 8132902]
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on Speech Perception Progress Report No.10. Speech Research Laboratory, Indiana University; 1984. Sizing up the Hoosier Mental Lexicon: measuring the familiarity of 20,000 words; p. 357-376.
- Nygaard, LC. Perceptual integration of linguistic and non-linguistic properties of speech. In: Pisoni, DB.; Remez, RE., editors. *The Handbook of Speech Perception*. Oxford, UK: Blackwell Publishing Ltd; 2008. p. 390-413.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Percept Psychophys*. 1998; 60(3):355–376. [PubMed: 9599989]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychol Sci*. 1994; 5(1):42–46. [PubMed: 21526138]
- Nygaard LC, Sommers MS, Pisoni DB. Effects of stimulus variability on perception and representation of spoken words in memory. *Percept Psychophys*. 1995; 57(7):989–1001. [PubMed: 8532502]
- Oyama S. A sensitive period for the acquisition of a non-native phonological system. *J Psycholinguist Res*. 1976; 5(3):261–283.
- Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. *J Exp Psychol Learn Mem Cogn*. 1993; 19(2):309–328. [PubMed: 8454963]
- Peters, RW. Joint Project Report No 56. Vol. 56. US Naval School of Aviation Medicine; Pensacola, FL: 1955. The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise; p. 1-9.
- Pisoni, DB. Some thoughts on “normalization” in speech perception. In: Johnson, K.; Mullennix, JW., editors. *Talker Variability in Speech Processing*. San Diego, CA: Academic Press; 1997. p. 9-32.
- Pisoni DB. Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception. *Ear Hear*. 2000; 21(1):70–78. [PubMed: 10708075]
- Pisoni, DB. *WordFam: Rating Word Familiarity in English*. Bloomington, IN: Indiana University; 2007.
- Pisoni DBM. Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear Hear*. 2003; 24(1 Suppl):106S–120S. [PubMed: 12612485]
- Pisoni DB, Geers AE. Working memory in deaf children with cochlear implants: correlations between digit span and measures of spoken language processing. *Ann Otol Rhinol Laryngol Suppl*. 2000; 185:92–93. [PubMed: 11141023]
- Pisoni DB, Kronenberger WG, Roman AS, Geers AE. Measures of digit span and verbal rehearsal speed in deaf children after more than 10 years of cochlear implantation. *Ear Hear*. 2011; 32(1, Suppl):60S–74S. [PubMed: 21832890]
- Pisoni DB, Nusbaum HC, Luce PA, Slowiaczek LM. Speech perception, word recognition and the structure of the lexicon. *Speech Commun*. 1985; 4(1–3):75–95. [PubMed: 23226910]
- Plomp R. Noise, amplification, and compression: considerations of three main issues in hearing aid design. *Ear Hear*. 1994; 15(1):2–12. [PubMed: 8194676]
- Pollack I. Auditory informational masking. *J Acoust Soc Am*. 1975; 57:S5.
- Ptacek PH, Sander EK. Age recognition from voice. *J Speech Hear Res*. 1966; 9(2):273–277. [PubMed: 5925532]
- Richards VM, Zeng T. Informational masking in profile analysis: comparing ideal and human observers. *J Assoc Res Otolaryngol*. 2001; 2(3):189–198. [PubMed: 11669393]
- Rogers CL, Lister JJ, Febo DM, Besing JM, Abrams HB. Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl Psycholinguist*. 2006; 27(3):465–485.
- Roth, R.; Isquith, P.; Gioia, G. BRIEF-A: Behavioral Rating Inventory of Executive Function – Adult Version: Professional Manual. Lutz, FL: Psychological Assessment Resources; 2005.
- Salthouse TA. The processing-speed theory of adult age differences in cognition. *Psychol Rev*. 1996; 103(3):403–428. [PubMed: 8759042]

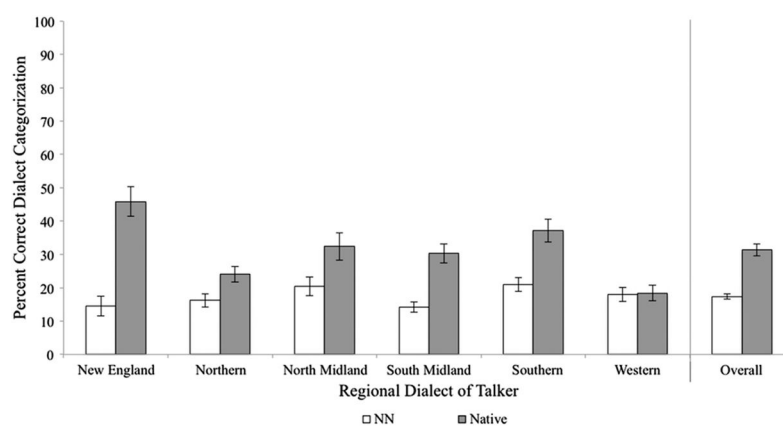
- Samuel, AG. The role of the lexicon in speech perception. In: Schwab, EC.; Nusbaum, HC., editors. Perception of Speech and Visual Form: Theoretical Issues, Models, and Research. Vol. 1. New York, NY: Academic Press; 1986. p. 89-111.
- Segalowitz, N. Individual differences in second language acquisition. In: de Groot, A.; Kroll, JF., editors. Tutorials in bilingualism: Psycholinguistic perspectives. Hillsdale, NJ: Lawrence Erlbaum Associates; 1997. p. 85-112.
- Stenfelt S, Rönnberg J. The signal-cognition interface: interactions between degraded auditory signals and cognitive processes. *Scand J Psychol.* 2009; 50(5):385–393. [PubMed: 19778386]
- Studebaker GA. A “rationalized” arcsine transform. *J Speech Hear Res.* 1985; 28(3):455–462. [PubMed: 4046587]
- Takata Y, Nábělek AK. English consonant recognition in noise and in reverberation by Japanese and American listeners. *J Acoust Soc Am.* 1990; 88(2):663–666. [PubMed: 2212289]
- Tamati TN, Gilbert JL, Pisoni DB. Some factors underlying individual differences in speech recognition on PRESTO: a first report. *J Am Acad Audiol.* 2013; 24(7):616–634. [PubMed: 24047949]
- van Bezooijen R, Gooskens C. Identification of language varieties: The contribution of different linguistic levels. *J Lang Soc Psychol.* 1999; 18(1):31–48.
- Van Engen KJ, Bradlow AR. Sentence recognition in native- and foreign-language multi-talker background noise. *J Acoust Soc Am.* 2007; 121(1):519–526. [PubMed: 17297805]
- Van Lancker D, Kreiman J, Emmorey K. Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices. *J Phonetics.* 1985a; 13:19–38.
- Van Lancker D, Kreiman J, Wickens T. Familiar voice recognition: Patterns and parameters: Part II. Recognition of rate-altered voices. *J Phonetics.* 1985b; 13:39–52.
- van Rooij JCGM, Plomp R. Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *J Acoust Soc Am.* 1990; 88(6):2611–2624. [PubMed: 2283434]
- van Wijngaarden SJ, Steeneken HJ, Houtgast T. Quantifying the intelligibility of speech in noise for non-native listeners. *J Acoust Soc Am.* 2002; 111(4):1906–1916. [PubMed: 12002873]
- von Hapsburg D, Champlin CA, Shetty SR. Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners. *J Am Acad Audiol.* 2004; 15(1):88–98. [PubMed: 15030104]
- von Kriegstein K, Smith DRR, Patterson RD, Kiebel SJ, Griffiths TD. How the human brain recognizes speech in the context of changing speakers. *J Neurosci.* 2010; 30(2):629–638. [PubMed: 20071527]
- Wightman FL, Kistler DJ, O’Byrne A. Individual differences and age effects in a dichotic informational masking paradigm. *J Acoust Soc Am.* 2010; 128(1):270–279. [PubMed: 20649222]



**Figure 1.** Individual percent-correct keyword recognition accuracy scores on PRESTO sentences averaged over all four SNRs for non-native listeners (NN) and native (Native) listeners reproduced from Gilbert et al (2013). Non-native listeners are represented by open bars, and native listeners are represented by gray bars. Horizontal bars indicate mean accuracy for each listener group.

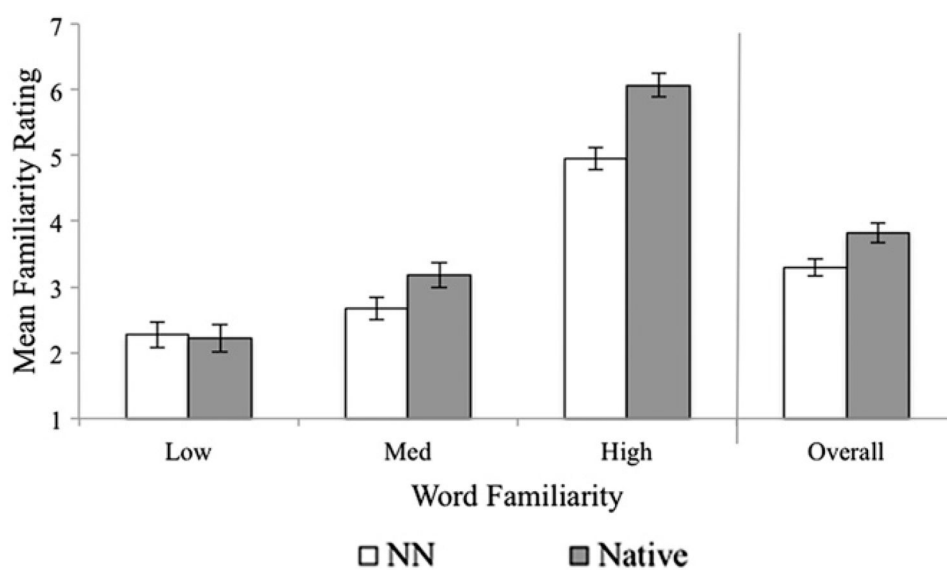


**Figure 2.** Mean percent-correct keyword recognition accuracy on PRESTO (dashed lines) and HINT (solid lines) sentences by SNR for the non-native (NN; black lines, circles) and native listeners (Native; gray lines, triangles). Error bars are standard error.

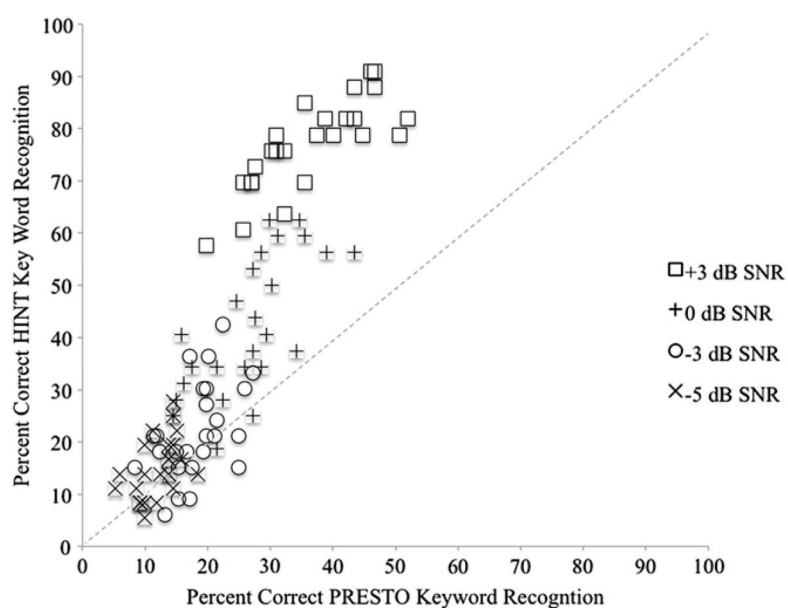


**Figure 3.** Mean percent-correct categorization accuracy on the Regional Dialect Categorization task across six American English regional dialects and overall for non-native (NN; open bars) listeners and native listeners (Native; gray bars). Error bars are standard error.





**Figure 4.** Mean familiarity ratings for low-, medium-, and high-familiarity words, and overall, on the WordFam questionnaire for non-native (NN; open bars) and native listeners (Native; gray bars). Error bars are standard error.



**Figure 5.** Individual performance (percent-correct keyword recognition accuracy) on PRESTO (x-axis) and HINT (y-axis) at all four SNRs for each of the 25 non-native listeners.

Table 1

Summary of Mean Keyword Recognition Accuracy Scores (%) for Both Listener Groups and Mean Difference Scores (%) between the Non-native and Native Groups on PRESTO and HINT, at Each SNR and Overall across All Trials and SNRs

SNR	PRESTO				HINT			
	Raw Scores (%)		Difference Scores (%)		Raw Scores (%)		Difference Scores (%)	
	Non-native	Native	Native – Non-native		Non-native	Native	Native – Non-native	
+3 dB	36.5	80.5	44.0		77.0	94.9		18.0
0 dB	26.7	64.5	37.7		42.3	79.8		37.5
–3 dB	18.3	46.6	28.3		22.2	44.7		22.6
–5 dB	11.9	30.6	18.7		14.8	24.1		9.3
Overall	23.2	55.5	32.4		38.5	59.9		21.4

**Table 2**

Results of Bivariate Correlations between Mean Overall PRESTO Keyword Recognition Scores and BRIEF-A Measures (One-Tailed) for Non-native Listeners

BRIEF-A	PRESTO Score		
	+3 dB SNR	0 dB SNR	Overall
GEC	$r = -0.47, p = 0.009$	$r = -0.60, p = 0.001$	$r = -0.49, p = 0.006$
MI	$r = -0.42, p = 0.019$	$r = -0.52, p = 0.004$	$r = -0.42, p = 0.018$
Self-Monitoring	$r = -0.35, p = 0.046$	$r = -0.50, p = 0.006$	$r = -0.38, p = 0.030$
Initiate	$r = -0.34, p = 0.047$	$r = -0.46, p = 0.010$	$r = -0.35, p = 0.043$
Working Memory	$r = -0.42, p = 0.018$	$r = -0.50, p = 0.006$	$r = -0.42, p = 0.019$
Plan/Organize	$r = -0.39, p = 0.026$	$r = -0.43, p = 0.015$	$r = -0.38, p = 0.030$
Task Monitoring	$r = -0.29, p = 0.082$	$r = -0.39, p = 0.029$	$r = -0.33, p = 0.052$
Organization of Materials	$r = -0.33, p = 0.053$	$r = -0.44, p = 0.013$	$r = -0.32, p = 0.058$
BRI	$r = -0.48, p = 0.008$	$r = -0.63, p < 0.001$	$r = -0.53, p = 0.003$
Shift	$r = -0.34, p = 0.048$	$r = -0.43, p = 0.017$	$r = -0.35, p = 0.043$
Inhibit	$r = -0.49, p = 0.007$	$r = -0.63, p < 0.001$	$r = -0.52, p = 0.004$
Emotional Control	$r = -0.43, p = 0.016$	$r = -0.57, p = 0.002$	$r = -0.51, p = 0.005$